Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Perceived Order in Different Sense Modalities: I. J. HIRSH AND C. E. SHERRICK, JR. 423
The Prediction of Free Recall from Word Association Measures: E. Z. ROTHKOPF AND E. U. COKE 433
Conditional Response Probability in a T Maze: R. S. WITTE 439
An Application of Stimulus Sampling Theory to Summated Generalization: T. S. CARTERETTE 448
Recognition of Complex Visual Stimuli as a Function of Training with Abstracted Patterns: M. H. MARX, W. W. MURPHY, AND A. J. BROWNSTEIN 456
Partial Reinforcement Effects (PREs) and Inverse PREs Determined by Position of a Non-rewarded Block of Responses: G. D. JENSEN 461
The Relationship between Kinesthetic Satiation and Inhibition in Rotary Pursuit Performance: R. S. LIPMAN AND H. H. SPITZ 468
Conflict and the Orientation Reaction: D. E. BERLYNE 476
Some Relations between the Intensive Properties of the Consummatory Response and Reinforcement: G. COLLIER, F. A. KNARR, AND M. H. MARX 484
Free Versus Serial Recall: N. C. WAUGH 496
The Effects of Amount of Reward, Requirement, and Several Related Probabilities on Human Performance: S. G. ZIPF 503
Perseveration as a Function of Degree of Learning and Percentage of Reinforcement in Card Sorting: A. ERLEBACHER AND E. J. ARCHER 510
Short-Term Retention as a Function of Average Storage Load and Average Load Reduction: L. S. REID, K. E. LLOYD, H. R. BRACKETT, AND W. F. HAWKINS 518
The Effects of Induced Muscle Tension during Tracking on Level of Activation and on Performance: L. R. PINNEO 523
Replication Report: Two Failures to Reproduce Effects of Anxiety on Eyelid Conditioning: M. S. KING, G. A. KIMBLE, J. GORMAN, AND R. A. KING 532
Supplementary Report: Effects of Instructions on the Perception of Multiple Targets: H. W. KARN AND L. W. GREGG 533

American Psychological Association

Vol. 62, No. 5 November 1961

Important Editorial Announcements on Cover 4

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*. When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1961 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 62, NO. 5 NOVEMBER 1961

PERCEIVED ORDER IN DIFFERENT SENSE MODALITIES[1]

IRA J. HIRSH AND CARL E. SHERRICK, JR.

Central Institute for the Deaf, St. Louis, Missouri

Perceptual discrimination and recognition require, as different stimuli, patterns that are formed by intensitive and qualitative change over space or time. In vision, patterning in space appears to be of primary importance, while in hearing the important dimension for patterning appears to be time (Hirsh, 1952; Piéron, 1952). Discrimination of these patterns requires a certain resolving power in the observer. In visual space the resolving power that appears to be crucial is visual acuity or the minimal separation between two lines or dots for perception of two, as it is also for space laid out on the skin. Spatial resolving power has also been studied in hearing, where it takes the form of the minimum separable angle between two sound sources for differential localization.

For patterns that are articulated in time, such as movement in vision or touch, and melody in hearing, a temporal resolving power is required. Its measurement, however, is complicated by the unidimensional nature of time. Figures and forms in visual or tactual space may be recognized and retain certain form characteristics even when they are reversed, but temporal patterns are quite different when they are reversed in time. Temporal resolving power must, therefore, be broken down into at least two measures: one, analogous to spatial acuity, is the minimum time that can separate two events so that they are just barely nonsimultaneous or successive; the other, given that two events are perceived as successive, requires a judgment of temporal order to be made.

The requirements for separating questions of successiveness from those of successive order have not only to do with experimental method, but also with the kinds of stimuli used. To investigate the limits of the ability to distinguish successive from simultaneous events, we need only a pair or a series of events about which we can ask *O* whether there were one or two or whether the series consisted of separate discrete events or one continuous event. When we investigate the ability to recognize temporal order, on the other hand, the two events in question must be distinctive so that *O* has some way of labeling

[1] This work was supported by a grant (G-4457) from the National Science Foundation to the Central Institute for the Deaf.

them differently and then of using these labels to identify which occurred first or second. Two identical stimuli that follow each other in time may be used to distinguish successiveness from simultaneity, but the question of successive order is not appropriate unless the two stimuli can be identified separately.

Simultaneity vs. successiveness.—One of the most extensive investigations having to do with successiveness was reported by Exner in 1875. He set for himself the task of determining how close together in time two stimuli may be and yet be perceived as not simultaneous (*ungleichzeitig*). Unlike more recent *Es*, Exner (1875) explored this temporal acuity in different sense modalities. In vision, he found that two brief flashes of light appeared to be nonsimultaneous when they were separated by 44 msec. or more. He reported certain variations in the result depending upon whether the stimulation was central or peripheral and upon whether both flashes stimulated the same place or different places. There were some conditions in which he reported apparent movement and in these cases he found that movement could be distinguished from simultaneity when the temporal separations were between 15 and 20 msec. He regarded movement, however, as a kind of artifact and relied more upon his 44-msec. value as descriptive of the temporal resolving power of the eye.

In hearing, using the successive spokes of a Savart wheel, he determined that the minimum temporal separation between successive clicks that would give rise to a judgment of successiveness was about 2 msec. The resolving power of the ear was clearly superior. When two clicks were delivered separately, one to each ear, the minimum temporal separation for successiveness was as large as 64 msec. Such a large difference probably has less to do with the two ears than with different experimental tasks, as will be noted presently.

Exner (1875) also reported some experiments on the minimum temporal separation between stimuli, one of which was presented to one sense modality and the other to another. The values are more like the large times of the visual temporal resolution or the one associated with stimulating the two ears separately, and furthermore they are not symmetrical; that is, in cases of intermodal stimulation, the value of temporal separation required for successiveness depended upon which of the two particular sense modalities was stimulated first.

Now Exner (1875) pointed out that there are four conditions under which successiveness versus simultaneity may be studied: (*a*) the stimuli could affect the same sensory elements, (*b*) they could affect different sensory elements in the same sense modality, (*c*) they could affect corresponding elements in paired sense organs, and (*d*) they could affect different sense modalities. More recently, Piéron (1952) and Fraisse (1957) have pointed out that the perception of simultaneity is easiest to obtain when stimuli occur at the same place and are presented to the same sense modality. Indeed, Fraisse (1957, p. 110) has indicated an order of difficulty among three classes of conditions which correspond to the first, second, and fourth conditions of Exner given above.

Particular values for the temporal acuity of different sense modalities based upon the distinction between successiveness and simultaneity have been given by many authors. Exner's (1875) values for vision and for hearing agree reasonably well with more recent data on visual flicker (Bartley, 1958, pp. 120–121), and on auditory successiveness (Miller & Taylor, 1948; Wallach, Newman, & Rosenzweig, 1949). Values of temporal acuity for the skin are less well established (Geldard, 1940). There is some evidence that different *Es* have used quite dissimilar criteria for tactual fusion. It is clear that this simple temporal acuity varies with the sense modality used and may indeed be ascribed to the resolving power of peripheral mechanisms.

As Exner's (1875) values for simple successiveness have been reported in secondary sources through the years (James, 1890, pp. 614–615; Ladd, 1892, pp. 364–366; and Troland, 1929, p. 399), they have been lumped together with the results of his other experiments involving stimuli presented to the two ears separately, to two parts of the retina, and to two different sense modalities. Fraisse (1957) comments that these latter cases of heterogeneous stimulation represent conditions in which perception of simultaneity is difficult to obtain. Indeed, although Exner (1875) is not very explicit about the particular instructions that he gave to his *Os*, it is doubtful that he asked them exactly the same questions in cases of homogeneous and heterogeneous stimulation. He does comment that he did not note a single case in which the time difference required for the perception of nonsimultaneity was different from that required to perceive which of the two stimuli came first. But, as has already been pointed out, the question of which came first could not

even have been asked in cases where the successive flashes of light stimulated the same place on the retina or where two Savart-wheel clicks followed each other. Those cases that Fraisse (1957) regards as difficult for obtaining judgments of simultaneity are the very ones in which the question of temporal order may be asked because the successive stimuli are different in quality, in place, or in sense modality. One of the conclusions of the present paper will be that the important distinction here is not between pairs of stimuli delivered to one sense modality as opposed to pairs delivered to separate sense modalities, as Fraisse (1957) and Piéron (1952) have suggested, but rather the more basic distinction is between experiments that ask about successiveness as opposed to simultaneity and those that require the judgment of temporal order. This conclusion will be justified if one can return to conditions of stimulation in one sense modality alone with differentiable stimuli whose order must be judged correctly and in which the results yield numbers more like those derived from heterogeneous stimulation where one has little choice but to ask for judgments of order.

Auditory perception of temporal order.—If two stimuli are presented to the same sense modality, are identical in quality and intensity, and occur in the same place, then one cannot ask about order of occurrence because there is no way for *O* to label separately each of the two stimuli. Here one is limited to asking whether there appear to be one or two. In a previous experiment in hearing (Hirsh, 1959), various pairs of sounds were manipulated with respect to time of occurrence, and in each pair the individual sounds were made different from each other with respect to frequency (if they were tones), spectrum (if they were bands of noise), and duration (if the pair consisted of both a brief and a prolonged sound). Under these circumstances, we could ask *O* which sound came first—the higher or lower pitch, the longer or shorter sound, the tone or noise, etc. The surprising result from these various combinations of sounds was that, whereas a 2-msec. separation is sufficient to yield nonsimultaneity, the time intervening between the two sounds that was required for judgments that were 75% correct was of the order of 20 msec. and was independent of the kinds of sounds used. The generality of this finding within the auditory domain suggested the present series of experiments whose aim it was to extend these observations to other sense modalities and also to pairs of stimuli that involved more than one sense modality.

Experiment I: Visual Order

The question for Exp. I was: how much time must intervene between two visual stimuli for *O* to report correctly the order in which they occur? The two stimuli were identical except that they appeared at different places—at two points along either a horizontal or a vertical line. The *O* was asked to report on whether the left or right or the upper or lower light came first; he could not say "simultaneous."

Procedure

Flashes of light were generated by electrical pulses of 5 msec. duration that fed an amplifier circuit in the output of which was connected a Sylvania R-1131C glow modulator tube. This glow tube was mounted in a box, its end facing *O*, behind a ground Plexiglas screen. A homogeneous illumination was achieved over a circular area of 0.5 cm. diameter. Level of illumination was set at 30 db. above absolute threshold. The pulses were generated by pulse generators (Tektronix Model 160) which in turn were driven by a single waveform generator (Tektronix Model 161) in such a way that the time interval between the two pulses could be manipulated with an accuracy of ± 1 μsec. The sequence of two pulses was repeated once per second.

The visual targets were placed on a table and, for the horizontal displays, were spaced 18, 36, or 72 cm. apart. With *O* 2 m. away, these separations corresponded to visual angles of 5°, 10°, and 20°. When the visual targets were disposed vertically, they were separated by 18 cm. The three horizontal distances were used with a red fixation point in the center so that the two white lights whose order was to be judged were to the left and to the right of center. In addition, an 18-cm. horizontal separation was used with the fixation point 18 cm. to the left of the left target and also 18 cm. to the right of the right target. The 18-cm. vertical separation was used with the fixation point both 18 cm. to the left of the vertical line and also 18 cm. to the right. These various conditions were employed to investigate the dependence of judgments of visual order upon the region of the visual field stimulated and also the spatial direction of the change.

One particular spatial arrangement was used in each experimental session. During each session 10 presentations of 7 different

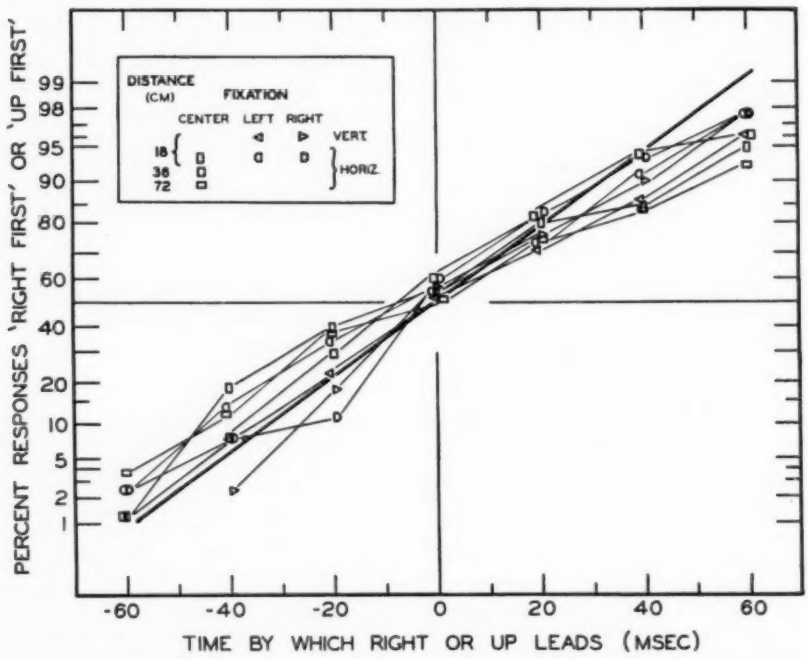

FIG. 1. Visual temporal order. Average percentage of responses "right first" or "upper first" as a function of the time by which the right or upper light preceded the left or lower light in milliseconds. (Parameters, indicated by different symbols as defined in the key, represent different spatial arrangements of the two lights. The heavy straight line comes from data on auditory perception of temporal order [Hirsh, 1959].)

temporal intervals were randomly ordered. The intervals ranged from 60 msec. with the left or lower light leading, to 60 msec. with the right or upper light leading, in 20-msec. steps. Four trained *O*s, between 18 and 30 yr. old, were tested singly. All sessions were repeated a second time with a different order of conditions for each *O* so that the results are based upon 20 presentations of each of the 7 temporal intervals under each of the 7 spatial arrangements.

Results

The average percentage of responses "right first" for horizontal displays and "upper first" for vertical displays is plotted in Fig. 1 as a function of the time by which the right or upper light leads the left or lower light. (Negative values on the abscissa thus represent the time by which the right or upper light lagged.) The different symbols denote the different distances and spatial arrangements of the visual targets. On the basis of a graphic inspection alone it appears that there are not systematic differences related to different spatial arrangements. The heavy line in Fig. 1 is the line that represents judgments of temporal order for auditory stimuli, being a summary of the several combinations of sounds used in a previous study (Hirsh, 1959). This function, which describes the auditory analog of this experiment, may be slightly steeper than a line that would be drawn through the average of the present visual data, but there is not a great difference

between the two, certainly not as much difference as one would anticipate on the basis of the relative speeds with which visual and auditory systems are supposed to act.

Results for mere successiveness (as opposed to fusion or simultaneity) reveal larger separation times for vision than for hearing. When *O* is asked to report on the order in which successive events occur (rather than on mere successiveness), however, the differences between these two sense modalities are small, if indeed they are real.

Experiment II: Auditory Order and Space

Experiment I differed from the previous studies on hearing (Hirsh, 1959) in that in the auditory experiment the two events whose order was to be judged differed from each other with respect to qualities like pitch, duration, etc., while in Exp. I, the two events whose order were to be judged were qualitatively the same visual stimulus but occurred at two different places. In spite of the relatively good agreement between the visual and auditory results, we were interested to know whether the auditory results would obtain if the two auditory stimuli whose order was to be judged were qualitatively the same but occurred at two different places.

Procedure

Pairs of pulses of 0.1 msec. duration were produced by the same system of pulse and timing generators that were used in Exp. I. These pulses were led separately to electrical simple tuned circuits. The resonant frequency of these tuned circuits was varied through the use of different fixed capacitors and the damping was also varied through the use of a variable resistor.

Three different conditions of stimulation were employed in order to make the two sounds whose order was to be judged (*a*) different with respect to pitch, (*b*) different with respect to which ear was stimulated, and (*c*) a combination of the first two.

The first of these conditions constitutes a repetition of one of the conditions in the previous auditory work (Hirsh, 1959). Here both pulses were presented to the same ear. One was passed through a circuit resonant at 666 cps and the other through a circuit resonant at 278 cps. The damping in both circuits was so adjusted that the amplitude of the damped oscillation that characterized the now-filtered pulse decreased to $1/e$ of its original value in 10 msec. Both pulses now sounded like brief xylophone notes of high and low pitch, which will be referred to hereafter as the "bing" and the "bong" respectively. Five new *O*s, trained young adults, were tested singly and asked to report which of the two pitches came first, the higher or the lower. Bings and bongs were separated by 10, 20, and 30 msec., with the bing leading one-half of the time and the bong leading the other half. A random series of 60 trials, 10 presentations at each of the six temporal separations, comprised an experimental session. No stimuli were presented with 0-msec. separation because a characteristic quality of simultaneity was always observed.

The second of the three conditions was intended to duplicate for the auditory system the spatial separation that characterized the visual experiment. In this case the two pulses were passed through the same high-frequency circuit (666 cps), one being led to the left ear and the other to the right ear. (We presume that this is Exner's [1875] two-ear condition.) The stimuli were now qualitatively the same, the difference between them being only the side on which the sound was heard. The same six temporal intervals were used. No stimuli were delivered with 0-msec. separation because we did not wish to confound the judgment of perceived order with the intracranial localization of a fused image. It should be clear that the temporal intervals used were all much longer than are involved in localization within the head or even fusion of the two stimuli into one sensation. Again a random series of 60 presentations, 10 at each of the six temporal separations, comprised an experimental session. Each of the 5 *O*s was asked to report which sound came first, the one on the right or the one on the left.

Finally, the two dimensions in which the auditory stimuli were different in the first two conditions were combined so that a high-frequency bing was presented to the right or left ear while the low-frequency bong was presented at the left or right ear, and *O* was asked to report which came first, the sound on the right or the sound on the left. His

response was still based upon location in space, even though the two events that occurred at the different places were also qualitatively different with respect to pitch.

Results

The results of these three subexperiments are shown in Fig. 2. The open circles show the percentage of judgments "higher first" as a function of the time by which the high-pitched bing leads the lower-pitched bong. The line connecting these open circles points to a finding that was reported on previously, namely that brief stimuli may in fact generate a function that is somewhat steeper (i.e., yields more precision) than the average for all auditory pairs. Judgments concerning which of two qualitatively identical bings comes first when they are delivered separately to the two ears are shown by the Xs. These points do not appear to deviate greatly from the straight line which again represents an average of results on perceived order for all sound pairs used in Hirsh's (1959) previous study. The closed circles show similar results for the condition in which the two events whose order is to be judged are different with respect to both place and quality. The lines connecting these points indicate a relatively broad range of indeterminacy around 50% between temporal separations of ±10 msec. Once beyond this range, however, the rate at which judgments change with increasing temporal separation appears to be greater than that for other conditions.

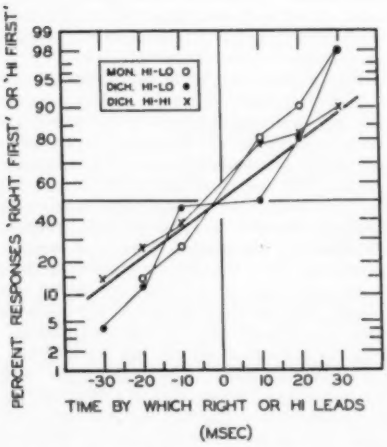

FIG. 2. Auditory temporal order. Average percentage of responses "right first" or "higher first" as a function of the time by which the right ear or the higher pitch preceded the left ear or lower pitch in milliseconds. (The open circles show results for monaural perception of order between two different pitches; the Xs show results for perception of order between the ears receiving the same pitch, and the filled circles show results for perception of order between two events that are different both with respect to ear and pitch. The straight line summarizes previous results in auditory perception [Hirsh, 1959].)

EXPERIMENT III: PERCEIVED ORDER AT TWO PLACES

So far we have demonstrated, among other things, that approximately the same amount of time is required to tell which of two stimuli came first when those two stimuli are placed on different sides, either in external space (visual) or in the head (auditory), no matter whether the visual or auditory modality is stimulated. To test the generality of this conclusion, we set up a similar experiment for the tactual system, using vibratory stimuli to the index fingertip of the right and left hands. The 5 *O*s were the same as those used in Exp. II but different from those used in Exp. I. We proposed to compare the percentage of responses "right first" as a function of the time by which the right led the left in three sense modalities: auditory, visual, and tactual. The first third of the experiment was already done and comprises the data on right versus left clicks in Exp. II (the X symbols in Fig. 2).

Procedure

The apparatus and procedure used for the visual part is exactly the same as discussed under Exp. I. Only one distance was used, 18 cm., the two target lights being on the same horizontal line with a red fixation point halfway between. In the same group of 5 *O*'s, each *O* was asked to say whether the right or left light came first as the temporal separation between the two lights was varied from 30 msec., with the left leading, to 30 msec. with the right leading, in 10-msec. steps.

The outputs of the same pulse generators were fed to a power amplifier whose output was delivered to two Goodmans V-47 vibrators. These somewhat resonant transducers damped the 1-msec. rectangular pulse in such a way that the mechanical output, when the transducer was loaded by a fingertip, resembled approximately a half-wave at 100 cps. The *O* sat with his arms resting on two tablet-arm chairs and his index fingertips resting lightly on the vibrators. His wrists could remain on the tablet so that only the weight of his finger was supported by the vibrator. His task was similar to that in the other experiments: pairs of stimuli were delivered to the two vibrators and he was asked which of the two fingers was stimulated first, the right or the left. Temporal intervals between the tactual stimuli were varied from 30 msec., with the left leading, to 30 msec. with the right leading, in 10-msec. steps.

Results

Several attempts were made in preliminary experiments to collect data on the temporal resolving power of the tactual system. The *O*s were asked to say which of the following paired locations were stimulated first: face or toe, finger or palm, finger or arm. Wide variations among *O*s were observed, partly because this particular range of temporal intervals serves the tactual sense in much the same way as temporal intervals of the order of microseconds serve auditory localization. For example, when the fingertip and the palm are stimulated successively, a time interval of from 20 to 40 msec. appears to be sufficient for *O* to say which spot was stimulated first, but for a temporal interval only somewhat smaller, for example about 15 or 20 msec., there is but one stimulus perceived and it is located neither at the fingertip nor at the palm, but at some place in between, depending upon the interval. Data from these preliminary studies are at best equivocal because of the wide variations among *O*s and will not be reported here.

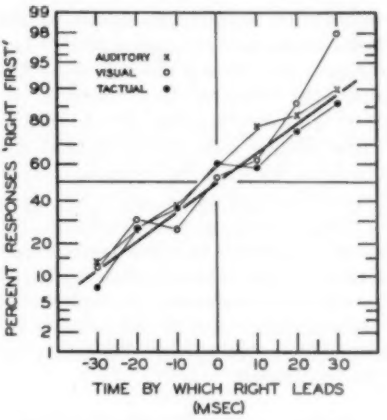

FIG. 3. Temporal order in space. Average percentage of responses "right first" as a function of the time by which the right stimulus precedes the left stimulus in milliseconds. (The different symbols represent three different sense modalities. The straight line summarizes previous results in auditory perception [Hirsh, 1959].)

The particular task described above is not so difficult if the places for tactual stimulation are on opposite sides of the body, namely the fingertips. Results averaged over the 5 *O*s are shown by the filled circles in Fig. 3. The visual experiment in which the targets were placed 9 cm. to the left or to the right of a fixation point yielded results shown by the open circles in Fig. 3. The Xs are taken directly from Fig. 2 and are used for comparison with tactual and visual data because that particular auditory task (which ear was stimulated first by the same quality click)

seems most nearly analogous to the tasks of the present experiment. The straight line drawn through Fig. 3 is again the function that described the early auditory results (Hirsh, 1959). These results indicate to us that the precision for judging temporal order between two places is quite the same in these three sense modalities. It seems reasonable to hypothesize that a single quantitative process is shared by all three modalities at some point. The experiments that follow may aid in deciding about the adequacy of this hypothesis.

Experiment IV: Perceived Order between Modalities

One further experiment seemed in order before we could establish the generality of data on perceived order and their independence of particular sense modalities. It has been noted that the classical literature on temporal resolving power did indeed treat the problem of perceived order of stimulation in two sense modalities (Exner, 1875). These data have led some to hypothesize that sensations in two sense departments cannot be so well organized in time as sensations in the same sense. But the data for the better temporal organization in single sense modalities have come from judgments of successiveness (as opposed to simultaneity) rather than from judgments of order. If our notion is correct, then the results indicate merely how much more time is required for judgments of perceived order when stimuli are presented to the same sense modality. Now we wished to test whether similar behavior would be observed when the two stimuli in each pair were delivered to different sense modalities.

Procedure

The apparatus has already been described. The additional variation here was that the output of one pulse generator was transduced by a glow modulator for a visual stimulus, while the second pulse was transduced by the vibrator for a tactual stimulus. By varying the temporal interval between the two pulses, we could vary the time by which the visual flash preceded or followed the tactual vibration. In this case, *O* was asked which came first, the tap or the flash.

When either the visual or tactual stimulus was to be paired with the auditory, then one of the pulses was filtered (666-cps bing) and transduced by an earphone.

The same 5 *O*s of Exp. II and III were tested singly. One session was devoted to a particular pair of modalities: either auditory-visual, visual-tactual, or auditory-tactual. Under each of these conditions, 70 stimuli were used, 10 presentations at each of 7 temporal intervals. All stimuli were maintained at a level 30 db. above threshold.

Results

Results are shown in Fig. 4. The different symbols refer to the three pairs of modalities. The open circles

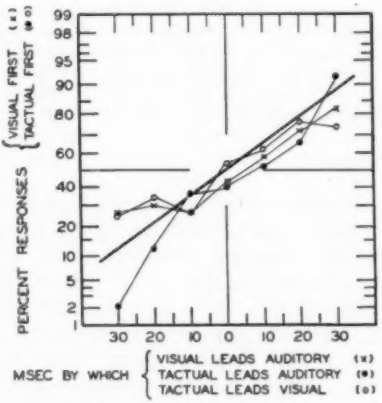

Fig. 4. Perceived order of stimuli affecting different sense modalities. (The ordinate shows the average percentage of responses "visual first" as a function of the time separating auditory and visual stimuli [Xs]; or "touch first" as a function of the time separating auditory and tactual stimuli [filled circles]; or "touch first" as a function of the time separating visual and tactual stimuli [open circles]. The straight line summarizes previous results in auditory perception [Hirsh, 1959].)

show the percentage of responses that were "touch first" as a function of the time by which the tactual stimulus led the visual stimulus. The negative numbers on the left side of the abscissa then would, for this particular set of symbols, indicate that the visual is leading the tactual stimulus. Similarly, the closed circles show the percentage of responses that were "touch first" as a function of the time by which the tactual led the auditory stimulus. Finally, the Xs show the percentage of responses that were "visual first" as a function of the time by which the visual flash led the auditory click. The straight line is again a summary of Hirsh's (1959) previous auditory experiments, and except for a few departures at the lower left, the points for the present three subexperiments seem not to deviate greatly from the same function.

Discussion

Differences among the auditory, visual, and tactual modalities with respect to temporal resolving power are well known. Judgments of successiveness as opposed to simultaneity, in the case of discrete stimuli (the temporal 2-point limen), or in the case of a long series of discrete stimuli (fusion), show that the visual system is relatively sluggish while the tactual system is less so, and the auditory system seems to be the fastest.

More than such simple resolving power is required when *O* is asked which of two sensations precedes the other. Temporal resolution for successiveness is, to be sure, a necessary prerequisite for this judgment of order, but it is not sufficient. The judgment of order requires that two pieces of information must be organized with respect to time, and the results of the present studies seem to indicate that it does not much matter where these two pieces of information come from; that is, they may come from different parts of the same sensory mechanism or they may even come from different sense modalities. The time required to insure, for example, judgments that are 75% correct is approximately 20 msec. It is much longer than the resolution times that give rise to successiveness in any of the modalities, although it comes close to that for the slowest system, namely the visual. It is so much longer than the resolution times of the several sense departments that it seems to point to some kind of time-organizing system that is both independent of and central to the sensory mechanisms. From this it would follow that the data for perceived order are similar even though different sense departments are used.

The much longer times obtained by Exner (1875) in his experiments can be accounted for partly by the psychophysical method he employed. At one point Exner described his method as one in which a judgment of correct order *10 times successively* was required to fix the value of time for perceived order. From the present data, such a stringent requirement would place the time values for all conditions at greater than 40 msec. This would be about the 98% point on the psychophysical functions in the figures shown. A word should be said about *O*'s task in the experiments. When *O* is asked to say that one perception equals another, or as in this case coincides with another, it is commonly found that he will show some variation in preference for the response "equal" over responses "greater," "less," etc. In the present experiments, it was found that *O*s in training complained repeatedly that the perceptions were simultaneous. They were forced to decide on one order or the other, however, and the result was as shown. This is not a new fact in psychophysics, but it has not been clearly stated heretofore in connection with the data on temporal order. Had the present writers chosen the method of average error as the technique for obtaining thresholds, it is undoubtedly true that the *SD* would have been considerably greater than 20 msec.

One of the exceptions to the function

that we have come to regard as a general rule, namely the straight line whose intercept at 75% is 20 msec., is the case of two brief auditory stimuli whose order is to be judged. In the previous auditory study (Hirsh, 1959), there was some evidence that the slopes of analogous functions, where clicks were paired with tones, were somewhat steeper than the average slope which was more typical of judged order for the onsets of tones and noises. This is also seen in the bing-bong auditory study represented in Fig. 2. Broadbent and Ladefoged (1959) have suggested that Hirsh's (1959) original results which indicated that perceived order could be judged correctly with as little as 20 msec. intervening between sounds, were a combined result of highly trained *O*s and the fact that when the temporal interval between two sounds that are already close together in time is changed, the overall quality of the two-sound complex is changed. They suggested that it was on this basis that *O*s were making their judgments and this may, in fact, be the case, especially for brief stimuli where the slope of the function is even steeper. However, this kind of judgment of complex quality cannot be the case in the further experiments presented here. To be sure, the time intervals used in the visual experiment lay in a range where sometimes apparent movement was seen and there the judgment of "which came first" may have been deduced from "in what direction did the light move." When the different sense modalities were stimulated, however, there is no impression of movement; there is rather only a first and a second—and the same function seems to fit the data equally well.

Summary

A series of studies has been carried out in which it has been shown that the amount of time that must intervene between two events in order for *O* to report correctly (75% of the time) which of those two events preceded the other is approximately 20 msec. The function that relates percentage of judgments to the temporal interval between the events appears not to be different even though the two events are either two sounds of different quality, or two light flashes in different places, or two vibrations, one to each hand, or two sounds, one to each ear, or two different stimuli, one to each of two sense modalities.

Whereas the time between successive stimuli that is necessary for the stimuli to be perceived as successive rather than simultaneous may depend upon the particular sense modality employed, the temporal separation that is required for the judgment of perceived temporal order is much longer and is independent of the sense modality employed.

References

Bartley, S. H. *Principles of perception.* New York: Harper, 1958

Broadbent, D. E., & Ladefoged, P. Auditory perception of temporal order. *J. Acoust. Soc. Amer.*, 1959, **31**, 1539.

Exner, S. Experimentelle Untersuchnung der einfachsten psychischen Processe. *Pflüg. Arch. ges. Physiol.*, 1875, **11**, 403–432.

Fraisse, P. *Psychologie du temps.* Paris: Press Univer. France, 1957.

Geldard, F. A. Perception of mechanical vibration. *J. gen. Psychol.*, 1940, **22**, 243–269.

Hirsh, I. J. Certain temporal factors in audition. *Science*, 1952, **116**, 523.

Hirsh, I. J. Auditory perception of temporal order. *J. Acoust. Soc. Amer.*, 1959, **31**, 759–767.

James, W. *Principles of psychology.* Vol. 1. New York: Holt, 1890.

Ladd, G. T. *Outlines of physiological psychology.* New York: Scribner, 1892.

Miller, G. A., & Taylor, W. G. Perception of repeated bursts of noise. *J. Acoust. Soc. Amer.*, 1948, **20**, 171–182.

Piéron, H. *The sensations.* (Trans. by Pirenne and Abbott) London: J. Garnet Miller, 1952.

Troland, L. T. *Principles of psychophysiology.* Vol. 1. New York: Van Nostrand, 1929.

Wallach, H., Newman, E. B., & Rosenzweig, M. R. Precedence effect in sound localization. *Amer. J. Psychol.*, 1949, **62**, 315–336.

(Received September 1, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 433–438

THE PREDICTION OF FREE RECALL FROM WORD ASSOCIATION MEASURES

ERNST Z. ROTHKOPF AND ESTHER U. COKE

Bell Telephone Laboratories

This is a report of some observations on frequency and sequence effects in the free recall of individual items from a word list and on the relationship of these effects to data from word association norms. Such observations are not new in the literature. Sequence effects ("clustering") have been described by Jenkins and Russell (1952) and by Jenkins, Mink, and Russell (1958). Frequency of recall effects have been reported by Deese (1959a, 1959b) who showed that nonlist words which are common association responses for the words of a list tend to occur as extralist intrusions and that this effect is an increasing function of the average association frequencies between the words which compose the list and the intruding term. Deese (1959a, 1960) has also shown that more words were recalled from lists with high degree of intralist associations than from lists of low degree of intralist associations.

The studies referred to above have concerned themselves with overall list characteristics and not with the recall of the various individual words within a list although the first Deese study (1959b) indirectly bears on this problem. The prediction of recall, from word association norms, for the various individual words within a list is somewhat complicated by the fact that interassociations among list words are generally complex and that no satisfactory rational basis for weighting these interassociations is now available. The present experiment was intended as a test of a reasonably simple scheme for predicting the recall of individual words from word association norms. This scheme involves the use of the number of words in the list to which a given word, W, is normally given as an association response, as a predictor for the frequency of recall of W. It grew from the hypothesis (Bousfield, 1953) that free recall was in part dependent on cuing ("relatedness increment") among list components. The scheme was tested on a list composed of 99 words from the Kent-Rosanoff Word Association Test.

METHOD

Materials.—The list of stimulus words which were used in the free recall task consisted of 99 words from the 100-word Kent-Rosanoff list. One of the words, *cheese*, was omitted because procedural convenience required that the number of words in the lists be divisible by three. The Kent-Rosanoff words were chosen because recent word association norms based on large numbers of *S*s were available (Russell & Jenkins, 1954). Each stimulus word was typed in capitals on a separate 3 × 5 in. card.

Procedure and subjects.—Three *S*s were tested at the same time. The deck of stimulus cards was shuffled and split into three numerically equal subdecks. One of these was placed face down before each *S*. At *E*'s signals, *S*s inspected each card of their subdeck for 5 sec. When *S*s had finished viewing their subdeck, *E* reshuffled each subdeck, reassigned them, and inspection was continued. This procedure was repeated until each *S* had inspected each of the 99 words, once, for 5 sec.

Immediately following a 15-min. rest pause, *S*s were asked to write down all the words of the stimulus set which they could remember. A period of 5 min. was allowed for recall. The *S*s wrote the recalled words on 3 × 5 in. paper pads, one word to each page.

All *Ss* ($N = 28$) were volunteer eleventh and twelfth grade students who were paid for their services. All *Ss* were taking a college preparatory course. They had participated on the previous day in an identification learning experiment which did not involve words either as stimuli or as response terms.

Results

The number of words in the list for which a given word, W, is normally given as an association response will be referred to as W's *cue number* or N_C. N_C was determined by counting all stimulus words (except "cheese") for which W was given as an association response in the 1952 Minnesota norms of the Kent-Rosanoff Word Association Test (Russell & Jenkins, 1954). Only those association responses were considered which occurred either with a normative frequency of more than 11 or which were among the 10 most common association responses to the stimulus word, whichever yielded the greater number of different association responses. N_C was obtained separately for each word of the list. The frequency distribution for words of various N_C in the list is shown in Table 1 (for additional intralist association data for the 99 word K-R list see Rothkopf & Coke, 1961). The incidence of words for which $N_C = 0$ is 26. Frequency of incidence in general decreases as cue number (N_C) goes up.

TABLE 1

Incidence of Words with Various Cue Number (N_C) in the Kent-Rosanoff List (Minnesota 1952 Norms)[a]

Nc	Frequency	Nc	Frequency
0	26	9	1
1	23	10	2
2	18	11	3
3	11	12	1
4	3	13	0
5	2	14	0
6	1	15	1
7	3	16	1
8	3	>16	0

[a] The word "cheese" has been omitted.

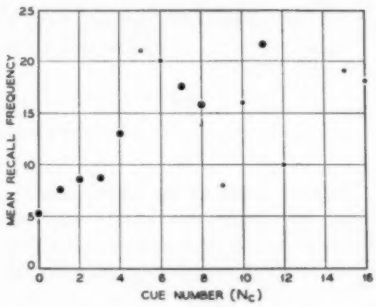

Fig. 1. Mean recall frequencies for words of various cue number. (The smaller datum points involve only three observations or less.)

Mean frequency of recall for words of the same N_C value was plotted against N_C in Fig. 1. As indicated before fewer observations were available for the high than for the low N_C values. The high N_C portions of the Fig. 1 plot may therefore be less reliable estimates of the frequency of recall function than the low N_C portions. For this reason homogeneity of variance throughout the N_C range was tested by the Bartlett technique prior to proceeding with an analysis of variance. The five N_C values for which only one observation was available were omitted from this analysis. The results of this test did not permit the rejection of the homogeneity hypothesis ($P > .3$). The analysis of variance, summarized in Table 2, shows that variance due to N_C was significantly greater than within-N_C variance ($F = 7.01$, $P < .01$). The linear component of between-N_C variance was significant ($F = 70.23$, $P < .01$). However, the remainder variance was also found to be significantly larger than the within variance although only at the .05 level ($F = 2.14$). An inspection of

the plot suggests that there may be some negative acceleration of the function. It seems reasonable to argue that there is a monotonic, increasing relationship between recall and N_C. The product-moment correlation between recall frequency and N_C was .62. A simple attempt to weight N_C by association frequencies did not improve the correlation with recall. Weighting of N_C was accomplished by summing the association frequencies with which W is given for its various stimulus words. These sums when correlated with recall yield an r of .54. The association frequency sums are, of course, highly correlated with N_C ($r = .84$).

In view of the fact that recall of words and their frequency of use are positively related (see Underwood & Schulz, 1960, pp. 56–57), a serious possibility exists that the function shown in Fig. 1 was the result of a correlation between N_C and frequency of use. In order to check this possibility correlations were computed between N_C, frequency of recall, and frequency of use as estimated from the Lorge-Thorndike semantic count (Thorndike & Lorge, 1944). The obtained r between N_C and use frequency was .41; r between use frequency and recall was .18. The partial correlation between frequency of recall and N_C with the effect of frequency of use held constant was .609. It must be concluded that while N_C and frequency of use are to some extent related, the effect of this relationship on the correlation between recall and N_C is very small.

TABLE 2

ANALYSIS OF VARIANCE OF FREQUENCY OF RECALL OF KENT-ROSANOFF WORDS AS A FUNCTION OF CUE NUMBER (N_C)

Source	*df*	*MS*	*F*
Between N_C	14	142.16	7.01*
linear	1	1424.27	70.23**
residual	13	43.54	2.14*
Within N_C	84	20.28	
Total	98		

* $P < .05$.
** $P < .01$.

If the effect of N_C on recall is to be accounted for in terms of a cuing mechanism, it should be expected that words will be recalled immediately after any one of the words for which they are normally given as association responses with frequencies which are in excess of chance. It should also be expected that the incidence of such sequences will increase with N_C even when the fact that words with high N_C are recalled more often than words with low N_C is taken into consideration. This was tested in the following manner. Whenever a given word followed, in free recall, immediately after a word for which it was an association response, this event was referred to as a *cue linkage*. In order to adjust for differences in recall frequency among words, the frequency of cue linkages in the recall protocols of all *S*s for every word was divided by the number of opportunities for cue linkage for that word. Opportunities for cue linkage were calculated by subtracting the number of times a word was recalled first (when no linkage could occur) from the total frequency with which that word was recalled. The mean ratio of cue linkage to opportunities for cue linkage is plotted against N_C in Fig. 2. The ratio of cue linkage to opportunities for cue linkage is an increasing function of N_C. This conclusion was supported by an analysis of variance, summarized in Table 3. Prior to performance of the analysis of variance, Bartlett's technique was used to test the hypothesis of homogeneity of variance throughout the N_C range. Again all

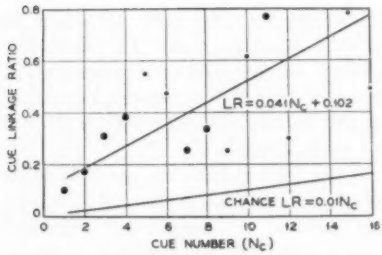

FIG. 2. Cue linkage ratio (cue linkage/opportunity for linkage) as a function of cue number. (The top line has been fitted to the data by the method of least squares. The lower line—chance function—is explained in the text.)

N_C values involving only one observation were not included. The homogeneity of variance hypothesis could not be rejected ($P > .9$). Table 3 shows a significant F ratio for between-N_C variance ($F = 5.61$, $P < .01$), and also indicates that the linear component of between-N_C variance was significant ($F = 50.24$, $P < .01$). The slope of the empirical function in Fig. 2 is significantly greater ($t = 5.037$, $df = 71$, $P < .01$) than the slope of the chance function which is also shown in Fig. 2. The chance function was based on the fact that words with greater N_C are more likely, by chance, in terms of the scoring criterion, to follow words for which they are

TABLE 3

ANALYSIS OF VARIANCE OF CUE LINKAGE: OPPORTUNITY FOR LINKAGE RATIO AS A FUNCTION OF CUE NUMBER (N_C)

Source	*df*	*MS*	*F*
Between N_C	13	.1811	5.60**
linear	1	1.6226	50.24**
residual	12	.0611	1.89
Within N_C	59	.0323	
Total	72[a]		

[a] There are 26 fewer *df* here than in Table 2 because Table 3 analysis did not include 26 words for which N_C was zero since for these words no cue linkage can occur.

** $P < .01$.

association responses. The chance function was calculated by Chance Function = $N_C/98$. The finding that cue linkage increases with N_C even when differences in recall frequencies and chance factors are compensated for, is consistent with the hypothesis that the positive relationship between N_C and recall depends on some cuing mechanism.

The following procedure was used to find whether cue linkage depended on association strength (frequency) between the cuing word and the link (cued) word. The association norms for each cue word were examined and all association responses which were stimulus words were noted and their frequencies recorded. It was determined for each cue linkage whether the association frequency of the linked word was above, equal to, or below the median association frequency of the relevant associations to the cue word. There were 265 cue linkages in 1000 recalled words. Of these 265, 178 were above median, 30 were equal to the median, and 57 were below the median. The 30 equal-to-the-median cases were equally divided between the above and the below median category and a binomial test was performed (Siegel, 1956, pp. 36–41). The two-tailed test was significant ($z = 7.31$, $P < .01$).

An estimate of the amount of organization which was imposed on sequence of recall through the associations among the stimulus words was obtained by comparing cue linkage scores for individual *S*s (hence referred to as AA, actually recalled words in actual order), with two different estimates of chance cue linkage scores. One chance cue linkage score, which will be referred to as AR (actually recalled words in random order), was obtained by taking the words actually recalled by

each *S*, arranging them in random order, and counting the cue linkages which occurred. Ten random arrangements were made for each *S*'s set of words. The mean cue linkage over the 10 random arrangements was used in the comparison tests. The second chance linkage estimate, referred to as RR (randomly chosen words in random order) was generated by randomly choosing words from the 99 word list. The number of words chosen was the number of words actually recalled by an *S* less the number of nonlist responses made by that *S*. That number of "nonlist" terms[1] was then added to the randomly chosen words and the resulting set was then arranged in random order. This process was repeated 10 times for each *S*'s actual recall frequency and the mean of these 10 values was used in the statistical analysis. Mean cue linkage score for the actual recall protocols (AA) was 9.50. Mean linkage for the AR and RR were 1.84 and .84, respectively. Because of marked heterogeneity of variance, Friedman's analysis of variance by ranks (Siegel, 1956, pp. 166–172) was used to compare the three conditions. The overall comparison as well as the AA:AR and AR:RR differences were highly significant ($P < .001$).

Discussion

The results of this experiment support the hypothesis that the frequency of recall of individual words within a list is at least in part determined by the number of list words (N_C) for which the word in question is normally given as an association response. This finding is consistent with previous observations reported by Deese (1959a) which linked average frequency of recall for lists with interassociations among list items.

Another finding of this experiment was that the frequency with which a word is recalled immediately after one of its "cues" is an increasing function of the number of "cues" for the word in question. This result, which held even when differences in recall frequency and chance factors were taken into account is consonant with earlier findings by Jenkins et al. (1958; Jenkins & Russell, 1952) that associates tend to occur consecutively in recall.

The comparisons among cue linkages in actual recall sequences (AA) and the computer generated chance linkage estimates (AR and RR) support the conclusions which were presented in the preceding paragraphs. The finding that AA data show more linkages than the generated AR sequences indicates that *S*s tend to recall associates in immediate sequence more often than can be expected by chance. The fact that the generated AR sequences result in greater cue linkage frequencies than the RR sequences implies that words which are actually recalled by *S*s are potentially more linkable, i.e., have higher N_C values, than words which are randomly chosen from the list. This conclusion follows from the data because both AR and RR involve random sequences which differ only in that the AR sequences are composed of actually recalled words while RR sequences were generated by random choices from the 99 word list.

Taken together, the results of this experiment are consistent with a single, cuing interpretation of both the frequency of recall and the sequencing phenomena which were the subject of this study. Such an interpretation has been proposed by Bousfield (1953) who calls the cuing effect "relatedness increment." A reasonable model for free recall which, when elaborated, may fit the present data and also the findings of Jenkins et al. (1958; Jenkins & Russell, 1952) and Deese (1959a, 1959b, 1960) may include the following elements. At any time in recall that word among the not yet

[1] The mean number of nonlist words in recall was three per *S*. The number of nonlist responses did not vary appreciably with amount of recall.

recalled words which has the greatest superthreshold habit strength is recalled next. When a word is recalled it increases the habit strength of all its association responses by amounts which are proportional to the association frequencies. In this way the position of the cued words in the hierarchy of habit strengths of words in the list may be temporarily changed. Cuing may also bring below-threshold habits temporarily to above-threshold strength. The failure to improve the prediction of free recall from cue number by weighting N_C by association frequency may be due to the fact that the words of the Kent-Rosanoff list differ in many other respects besides interassociations. It is also possible and perhaps likely that available word association norms do not provide adequate indices of interitem association strength and that new indicators need to be sought.

Summary

It was hypothesized that the frequency of recall of a given word W in a free recall situation depends directly on the number of cues which are provided by other words in the list. The cue number of word W was defined as the number of words in the learning list for which W is normally given as an association response in a word association test. The hypothesis was tested by exposing *Ss* ($N = 28$) to each of 99 words of the Kent-Rosanoff list for 5 sec. and testing for recall after 15 min.

The data supported the following conclusions: (*a*) The frequency with which a word was recalled was an increasing function of its cue number. (*b*) The frequency with which a word was recalled immediately after one of the words for which it was an association response (cue linkage) was an increasing function of its cue number even when the effects of recall frequency and chance differences due to cue number were taken into account. (*c*) Words which followed one of their cues in recall tended to be among those word associations which had association frequencies above the median association frequency of the association responses (list words only) to the cue word. These results were consistent with a cuing interpretation of the frequency and sequence of recall phenomena which were the subject of the present study.

References

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

Deese, J. Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, 1959, **5**, 305–312. (a)

Deese, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22. (b)

Deese, J. Frequency of usage and number of words in free recall: The role of association. *Psychol. Rep.*, 1960, **7**, 337–344.

Jenkins, J. J., Mink, W. D., & Russell, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127–136.

Jenkins, J. J., & Russell, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–821.

Rothkopf, E. Z., & Coke, E. U. Intralist association data for 99 words of the Kent-Rosanoff word list. *Psychol. Rep.*, 1961, **8**, 463–474.

Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. N8-ONR-66216)

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

Underwood, B. T., & Schulz, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Early publication received March 27, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 439–447

CONDITIONAL RESPONSE PROBABILITY IN A T MAZE[1]

ROBERT S. WITTE

San Jose State College

In a previous study (Witte, 1959), a trace-fluctuation model derived from statistical learning theory was applied to choice behavior under sequentially dependent reward schedules. An adequate test of the theory was hampered by the large number of *S*s with position preferences. The present study attempts to test the theory further by using a procedure designed to yield within-*S* comparisons, as well as to minimize position preferences.

When reinforcement is independent of *S*'s responses and occurs on every trial, the trace-fluctuation model generates the well-known "probability-matching" prediction for unconditional response probability. When, in addition, trials are massed, the model predicts that conditional response probability will be a direct function of the sequential dependencies between reinforcements. More specifically, the asymptotic probability of a response similar to the one reinforced on the previous trial, henceforth called a "repetition" response, following Anderson's (1960) usage, is

$$p_{ii}(\infty)=\frac{1}{N_e+N(1-j)^t}\{N_e[\pi+\theta(1-\pi)]+N(1-j)^t[\pi_{ii}+\theta\pi_{ii}(1-\pi_{ii})]\} \quad [1]$$

where $p_{ii}(\infty)$ is the asymptotic probability of a repetition response; N_e is the number of elements in the set of common stimulus elements; N is the number of "trace" elements in the set associated with an E_i reinforcing event; j is the rate at which the N trace elements become inactive; t is the size of the time interval since E_i occurred; π is the unconditional probability of E_i; θ is a sampling parameter; and π_{ii} is the conditional probability of E_i given E_i on the previous trial.

According to Equation 1, if the values of N_e and j are small, the asymptotic probability of a repetition response varies approximately from π_{ii} to π as a function of increases in t, the intertrial interval parameter. If the values of N_e and j are not small, asymptotic probability still varies with t, but the range of variation is curtailed in the direction of π.

Equation 1 refers only to the asymptotic probability of repetition responses following one previous reinforcing event. If a nonnegligible value of $(1-j)^t$, the quantity representing the stimulus trace, spans more than one intertrial interval, Equation 1 must be expanded to handle traces from more remote events. Such an expansion has been done (Witte, 1959), and henceforth it will be referred to as Equation 1a.

Equation 1a predicts the asymptotic probability of repetition responses as a function of the number of available traces, x, where x ranges in value from 0 to n. When predicted values are plotted against x, the form of the resulting function depends primarily on the relative value of reinforcement and the particular sequence of events.[2] Thus, for example, if $\pi_{ii}>\pi$, an increasing monotonic function occurs when the n traces originate from a sequence of n

[1] This investigation was conducted at Cornell College. It was supported in part by a research grant (M-2848A) from the National Institute of Mental Health, United States Public Health Service, and in part by a grant (G-7962) from the National Science Foundation. Paul Dude, Brian Klepinger, James Overly, Jenny Waychoff, and Robert Beckmen collected most of the data. The final draft of this paper was written during the tenure of a postdoctoral fellowship (MF-11219) from the National Institute of Mental Health, United States Public Health Service.

[2] In some cases, the form of the function also depends on the values of θ and N_e/N, but this complication is ignored for simplicity.

similar events ($E_iE_iE_iE_i\cdots$), but a decreasing cyclical function occurs when the n traces originate from a sequence of n alternating events ($E_jE_iE_jE_i\cdots$). Although too numerous for itemization, all possible combinations of reinforcement values and event sequences yield functions which are determined uniquely by Equation 1a.

Regardless of the value of x, Equation 1a generates essentially the same predictions as Equation 1 for the asymptotic probability of repetition responses as a function of t. To obtain a single, composite prediction from Equation 1a for a particular value of t, one can find the expected predicted value over all possible sequences of n events. As this procedure is repeated for larger values of t, the expected value approaches π in agreement with the trend predicted by Equation 1 for the special case where $x = 1$.

Several studies lend qualitative support to the predictions of the trace-fluctuation model for the massed-trial case. Both Engler (1958) and Anderson (1960) used human *S*s in a verbal conditioning design involving sequentially dependent rewards. In general, these *E*s found probability matching between conditional (and unconditional) response and reward probabilities. After analyzing heretofore unpublished data, Engler (1958) also reported that *S*s in the study by Hake and Hyman (1953) matched conditional response and reward probabilities.

Anderson (1960) tested two models from statistical learning theory, including one proposed by Burke and Estes (1957) which has many properties in common with the present theory. He found large and consistent terminal differences between predicted and obtained values when $\pi_{ii} < .40$. Anderson also suggested that neither model which he tested was able to account for the observed proportions of repetition responses following various sequences of reinforcing events.

The trace-fluctuation model was tested explicitly in a T maze study with rats (Witte, 1959). As trial spacing was increased from 8 sec. to 1 hr., the terminal proportions of repetition responses approached π in accordance with theoretical expectations. However, the quantitative fit between theory and data was not conspicuously successful for either the above analysis or a more detailed dealing with behavior following various reward sequences.

These discrepancies may have been caused by the selective elimination of *S*s with position preferences. The remainder of this paper describes an attempt to test the theory under conditions designed to minimize position preferences or, at least, to avoid the selective elimination of *S*s.

Method

Design.—The major independent variables were two symmetrical, sequentially dependent reward schedules and four intertrial intervals. Except for the 24-hr. interval between experimental days, all *S*s worked under a 12-sec. interval during the 360-trial "acquisition" series. Then, presumably after a behavioral asymptote had been reached, *S*s under a given reward schedule were divided at random into four groups. These groups underwent 50 "asymptotic" trials in each of four intertrial intervals in accordance with a systematically selected latin square design. The intervals were 12 sec., 32 sec., 3 min., and 30 min.

A correction procedure insured that each *S* was rewarded equally often on both sides of the T maze. Approximately three-fifths of the *S*s worked under an "alternation" reward schedule in which E_i followed E_i 10% of the time, while E_i followed E_j 90% of the time. The remaining *S*s worked under a "repetition" reward schedule in which E_i followed E_i 90% of the time, while E_i followed E_j 10% of the time. A reward schedule was predetermined separately for each *S* over an initial block of 60 trials and subsequent blocks of 50 trials. These schedules were constructed from tables of random numbers with the restriction that the sequence re-

flected, within a margin of one event, the unconditional and conditional frequencies specified in the experimental design.

Several features of the design merit additional comment. Previous findings (Witte, 1959) indicated that position preferences occurred most frequently among *Ss* under highly distributed trials and among *Ss* under alternation reward schedules. In an attempt to minimize the former effect, trials were massed during the acquisition series for all *Ss*, and to compensate for the latter effect, a majority of *Ss* were assigned to the alternation schedule. Notwithstanding these precautions, a large incidence of position preferences occurred among the alternation *Ss* during acquisition. Accordingly, only the repetition *Ss* were rotated through the four intertrial intervals of the asymptotic series.

Apparatus.—The apparatus, a simple T maze, has been described elsewhere (Witte, 1959). In the present study, a heavy cardboard partition, extending from the maze table to the ceiling of the experimental room, separated the two branches of the maze from a desk chair which straddled the stem of the maze. The partition had two $8\frac{1}{2} \times 18\frac{1}{2}$ in. horizontal apertures which permitted *E* to observe *Ss*' movements in the branches. Covered with light blue gelatin paper (#129, Rosco Lab.), the apertures functioned as one-way screens, since the only light sources in the room were three 100-w. bulbs on the branch side of the partition.

Procedure.—A $22\frac{1}{2}$-hr. food deprivation schedule was enforced throughout the experiment. All groups of *Ss* participated in a counterbalanced design yielding a mean deprivation of 20 hr. at running time.

After a 12-day training period, described elsewhere (Witte, 1959), trials in the T maze were begun. An additional 12 days elapsed before *Ss*' running times permitted the desired five trials per day. On each trial, after *S* had been placed in the startbox, the gates at the startbox and at the choice point were raised. When *S*'s entire body had entered a given branch alleyway, these gates were lowered, and one of the goalbox doors was unlocked according to a predetermined schedule. If *S* ran to the side containing the locked door, he was forced to return to the choice point and proceed to the other side. After *S* had eaten a .045-gm. food tablet (P. J. Noyes Co.), he was removed from the goalbox and placed in an individual detention cage for a prescribed period of time before receiving the next trial.

By definition, the intertrial interval began when *E* removed the lid of the goalbox, and it ended when *S* reached the choice point on the next trial. Since the time required by most *Ss* to get from the startbox to the choice point stabilized at about 2 sec., the intertrial interval could be specified within a small margin of error. Because of the correction procedure, the size of the *intratrial* interval was relatively variable, but it averaged about 8 sec.

A pilot study showed that a single *E* tended to be harried, and as a result, probably cued the location of the reward on some trials. In order to minimize inadvertent cues, *Es* worked in pairs. One *E* handled *Ss* and operated the gates in the maze stem, while the other *E*, sitting in the desk chair described above, recorded *Ss*' responses, controlled the goalbox doors, and timed the intertrial intervals. All *Es* were closely supervised, and special meetings were conducted to insure a uniform treatment.

Subjects.—The *Ss* were 64 naive hooded rats between 5 and 6 mo. of age at the beginning of the experiment. Three fatalities occurred during the experiment, and 6 *Ss* were discarded because they consistently refused to run in the maze.

Results

The response was defined as the first choice of *S* on each trial. Left and right turns at the choice point were designated as A_1 and A_2 responses; food rewards in the left and right goalboxes were identified with E_1 and E_2 events. Unless noted otherwise, all analyses were based on responses from Trials 2–5 of each experimental day, and terminal estimates of response probability for the acquisition series were derived from the last 20 days of the series.

Unconditional probabilities.—According to statistical learning theory (Estes, 1954), the terminal A_1-response proportions of individual *Ss* should approximate a binomial distribution about π. To test the group mean or π prediction, terminal estimates of .51 ($s_p = .046$) and .52 ($s_p = .014$) were obtained for the 31 alternation *Ss* and the 24 repetition *Ss* during the acquisition series, and an estimate of .48 ($s_p = .018$) was

obtained for the repetition *S*s during the 40-day asymptotic series. Obviously, none of these estimates deviated significantly from the predicted value of $\pi = .50$.

To test the binomial prediction, the expected distribution of A_1-response frequencies was determined for *S*s within a given group by referring to tables of the binomial distribution (Office of Chief of Ordnance, 1952), and the discrepancies between obtained and expected frequencies were evaluated by computing D, the statistic associated with the Kolmogorov-Smirnov test for goodness of fit (Siegel, 1956).

The terminal distribution of A_1 responses for the alternation *S*s during acquisition departed significantly ($D = .39$, $P < .01$) from binomiality. A further analysis revealed that nonbinomiality was caused primarily by the large incidence of *S*s with position preferences. More specifically, 22 of the alternation *S*s fell outside the 99% acceptance interval for .50. The stability of these preferences was indicated by their early appearance during the 80-day acquisition series. For example, over one-half of the preferences appeared during the first 30 days. It should be noted, however, that there was no consistent preference among the 22 *S*s for a particular side of the maze. This was also the case for all other subgroups of *S*s with position preferences in the present study.

Although the repetition *S*s also departed significantly ($D = .37$ and $.38$, $P < .01$) from binomiality during the terminal phase of acquisition and the first asymptotic rotation, they approximated ($.15 \leq D \leq .17$, $P > .20$) the predicted distribution during each of the last three asymptotic rotations. The two departures from binomiality by the repetition *S*s were due mainly to bimodality and skewness, rather than to a large incidence of position preferences. Thus, in the acquisition analysis, 3 *S*s barely exceeded the 99% acceptance interval for .50, their preferences appearing only during the final 20–30 days. In the asymptotic analyses, a total of 6 *S*s fell outside the acceptance interval during at least one rotation. Typically, these preferences lasted only for one rotation, and they were distributed evenly among the four rotations and the four intertrial intervals.

The π and binomial predictions were also applied to A_1 responses on first daily trials during the last 20 days of acquisition for both groups, and during the 40 asymptotic days for the repetition *S*s. In all three cases, although the resulting mean estimates deviated insignificantly from π, the distributions departed significantly ($.32 \leq D \leq .42$, $P < .01$) from binomiality. In every case, the departure from binomiality was caused primarily by the large number of position preferences: 17 for the alternation *S*s, and 15 and 12 for the repetition *S*s during the acquisition and asymptotic series. These results for the repetition *S*s, when compared with the previous paragraph, indicate less stereotypy during later daily trials.

Conditional probabilities.—In this section, measures of conditional response probability reflect the performance of all *S*s within a given group. Unless noted otherwise, it may be inferred that no appreciable difference existed between the overall measure and that for any subgroup of *S*s without position preferences.

In addition to minimizing position preferences, the acquisition series was introduced to produce a stable proportion of repetition responses before E manipulated the size of the inter-

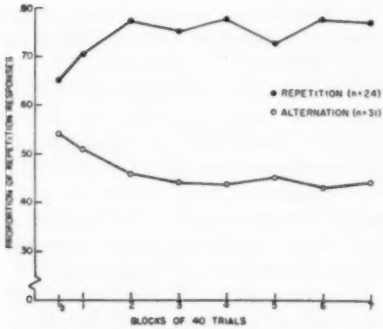

FIG. 1. Obtained proportion of repetition responses during acquisition.

trial interval. The acquisition curves for repetition responses are shown in Fig. 1. Both groups fell above the 50% level in the initial trials. Following rapid changes over Trials 21–80, i.e., Days 13–28, the alternation *Ss* approached a terminal value of .44 ($s_p = .016$), and the repetition *Ss* stabilized at .78 ($s_p = .013$). The curve for the subgroup of 9 alternation *Ss* without position preferences was lower by .03 than the alternation curve of Fig. 1. Because of the large incidence of position preferences among alternation *Ss*, no further results will be given for these *Ss*.

During the asymptotic series, as will be recalled, the repetition *Ss* spent 10 days under each of the four intertrial intervals. According to the theory, the asymptotic proportion of repetition responses for these *Ss* should approximate π as trial spacing increased. The obtained proportions, shown in Table 1, provide qualitative support for this prediction.

An analysis of variance was performed over the obtained proportions summarized in the first four intervals of Table 1. The effects associated with rotation period, order of presentation, and within-*S* interaction were not significant, while, not unexpectedly, the intertrial interval effect was significant ($F = 14.06$, $df = 3/6$, $P < .001$). A further analysis of the latter effect indicated that all paired comparisons, except the one between 12- and 32-sec. intervals, reached at least the .05 level of significance. A final analysis compared performance under the 12-sec. interval during the asymptotic series and during the terminal portion of the acquisition series. The resulting mean difference of .03 was not significant, as expected, assuming that a behavioral asymptote had been reached during the acquisition series.

Values of six parameters were required before the predicted proportions in Table 1 could be derived from Equation 1a. Three of these parameters, including the mean intratrial interval parameter, c, which appears in Equation 1a, were evaluated prior to an inspection of the repetition response data. Their values were $\pi_{ii} = .90$, $\pi = .50$, and $c = 8$.

The remaining parameters were estimated from the repetition response data after a simplifying assumption had been made. According to the theory, if no traces from previous reinforcing events are present on a given trial, only the set of common stimulus elements remains to be

TABLE 1

PROBABILITY OF REPETITION RESPONSES AS A FUNCTION OF INTERTRIAL INTERVAL: ASYMPTOTIC SERIES

Probability	Intertrial Interval				
	12 Sec.	32 Sec.	3 Min.	30 Min.	24 Hr.
Predicted	.75	.74	.73	.60	.52
Obtained	.75	.73	.66	.61	.53[a]
SE[b]	.019	.021	.022	.017	.016

[a] Based on first daily responses during the asymptotic series.
[b] Standard error of obtained probability.

sampled, and therefore, the asymptotic proportion of repetition responses on that trial should equal π (plus a small "correction" term). The proximity of π and .53, the obtained proportion of repetition responses on all first daily trials during the asymptotic series, implies that no traces spanned the 24-hr. interval between experimental days. This inference receives additional support from an estimate of .52 for first daily trials during the last 20 days of the acquisition series.

The simplifying assumption had several immediate consequences. First, the difference between observed proportions of repetition and alternation responses on first daily trials provided a sequential estimate of θ (Burke & Estes, 1957; Estes & Straughan, 1954). A sequential estimate of .04 for θ was obtained from the above proportion for the acquisition series, rather than its asymptotic counterpart, in order to avoid estimating parameters with to-be-predicted data.

Second, it could be inferred that not more than one trace was present on Trial 2 of each experimental day. Accordingly, Equation 1, instead of the more elaborate Equation 1a, was used to obtain estimates of .001 for j, the rate of decay, and .95 for N_c/N, the ratio of common elements to trace elements. These estimates represented solutions for a pair of expressions involving Equation 1 and two data points: the proportion of repetition responses on Trial 2 during the last 20 days of acquisition and the proportion on Trial 2 during the 10 asymptotic days which *S*s spent under the 30-min. intertrial interval.

Given the value of the intertrial interval, as well as the values of the six parameters, Equation 1a predicted the proportion of repetition responses on a particular daily trial as a function of each possible sequence of rewards. The number of possible reward sequences equaled 0, 1, 2, 4, and 8 for daily Trials 1–5, respectively.[3] To obtain a single, composite predicted value for any interval in Table 1, every prediction of Equation 1a was multiplied by the observed relative frequency with which its associated reward sequence had occurred, and these products were summed across all possible daily sequences.

A comparison of predicted and obtained proportions in Table 1 reveals a fairly large discrepancy at the 3-min. interval ($t = 3.0$, $P < .01$). Otherwise, the fit was excellent, particularly since only data from the 30-min. interval were used in estimating theoretical parameters.

Sequence probabilities.—An analysis of repetition responses as a function of daily reward sequences permits a more incisive, "fine-grained" test of Equation 1a. To facilitate this analysis, the immediately preceding reward, relative to which the repetition response was defined, is denoted by 1, regardless of whether it was an E_1 or E_2 event. More remote events are designated 1 or 2, depending on whether or not they differed from the referent event. Thus, for example, the symbol R(221) represents the proportion of repetition responses on Trial 4, given the referent event on Trial 3 and the alternative event on Trials 2 and 1. Note that the numbers in the above parentheses read from right to left. The present tabulational procedure differs from those used by Anderson (1960) and Nicks (1959) in that the data from each trial are being considered separately.

Table 2 shows the 10 most frequently occurring R values for the repetition *S*s, ranked according to predictions from Equation 1a. Except for an occasional reversal, nota-

[3] For example, the two possible reward sequences on Trial 3 were E_iE_i and E_jE_i.

TABLE 2

PROBABILITY OF REPETITION RESPONSES AS A FUNCTION OF REINFORCEMENT SEQUENCES: REPETITION *Ss*

Sequence[a]	*N*[b]	Blocks of 40 Trials										
		Acquisition							Asymptotic Interval			
		1	2	3	4	5	6	7	12 Sec.	32 Sec.	180 Sec.	1800 Sec.
R(1111)	160	.80	.84	.86	.90	.85	.89	.91	.88	.85	.73	.71
R(111)	184	.83	.85	.89	.85	.82	.88	.88	.88	.83	.70	.67
R(11)	213	.75	.87	.78	.87	.77	.86	.84	.85	.79	.77	.65
R(1)[c]	242	.65	.68	.73	.69	.66	.73	.70	.68	.67	.67	.57
R(2111)	19	.78	.76	.74	.82	.76	.74	.68	.67	.67	.62	.62
R(211)	21	.62	.62	.56	.59	.65	.50	.63	.39	.61	.41	.50
R(2211)	23	.68	.71	.50	.65	.53	.52	.47	.50	.25	.31	.62
R(21)	22	.46	.47	.40	.32	.50	.43	.37	.26	.33	.45	.36
R(221)	22	.55	.41	.22	.26	.35	.40	.32	.23	.31	.30	.35
R(2221)	22	.33	.26	.24	.35	.34	.20	.22	.07	.45	.26	.23
Summary *d*[d]		12	10	6	10	2	4	0	2	6	10	4

[a] Ranked from highest to lowest in terms of predicted values.
[b] Mean total frequency of occurrence per block of 40 trials.
[c] The theory predicts R(2111) > R(1) for the 180-sec. interval and R(2111) > R(211) > R(1) for the 1800-sec. interval.
[d] The letter *d* refers to the sum of the absolute discrepancies between predicted and obtained ranks, and it has a maximum value of 50.

bly at the fourth block of acquisition trials and at the 3-min. interval, the qualitative fit between predicted and obtained ranks improved progressively and was eventually quite good.

The large and consistent obtained differences between selected pairs of R values, e.g., R(1111) and R(2111), implies that *Ss*' behavior was influenced by the more remote daily

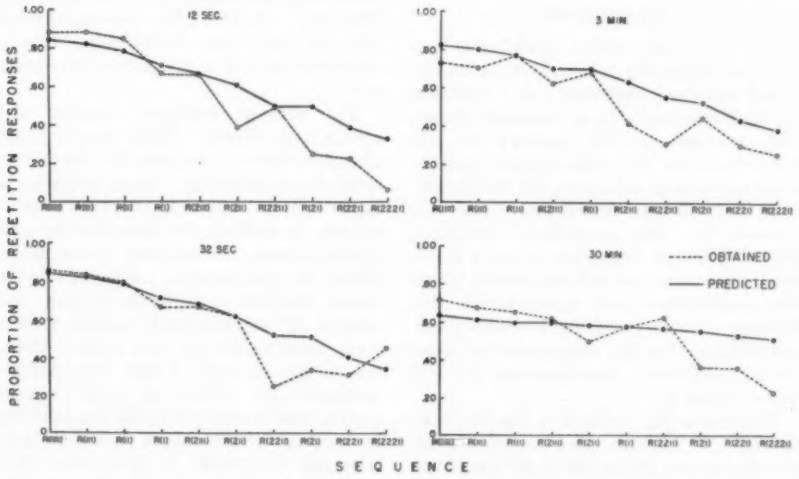

FIG. 2. Predicted and obtained R values during the four intertrial intervals of the asymptotic series.

rewards. With minor exceptions, however, the rank distributions for the four most frequently occurring R values in Table 2 failed to reflect Jarvik's (1951) negative recency effect. Instead, in agreement with the theory, repetition responses occurred more frequently as sequences of similar events increased in length.

The asymptotic R values for the repetition *Ss* were also compared with the set of exact predictions which had been computed previously to determine the expected proportions in Table 1. Inspection of Fig. 2 reveals that Equation 1a consistently overestimated R values for dissimilar events ($P < .01$), and underestimated R values for similar events, although the latter tendency was not significant, according to a binomial test. Generally speaking, the discrepancies were smaller for similar events than for the less reliable dissimilar events. The best fit between theory and data occurred at the 32-sec. interval, while the poorest fit occurred at the 3-min. interval, where the theory overestimated all R values except one.

Discussion

The trace-fluctuation model was formulated originally to account for conditional response probability as a function of trial spacing. In a previous study, and especially in the present one, its predictions for the trial-spacing variable were confirmed, primarily for the repetition *Ss*. When predictions were extracted for the conditional response probabilities of repetition *Ss* as a function of various reward sequences, both the qualitative and quantitative fits between theory and data were good, notwithstanding the appearance of some fairly consistent discrepancies in the quantitative fit.

Unfortunately, because of the large incidence of position preferences among the alternation *Ss*, an adequate test of the theory still is lacking for these *Ss*. During massed trials, substantially more position preferences appeared among the alternation *Ss* in the present study than among those in a previous study (Witte, 1959). Although the two studies are not strictly comparable because of differences in procedure and *Ss*, an important factor may have been the 4-sec. increase in trial spacing for the present study. This conjecture is strengthened by the fact that in the earlier study, position preferences appeared among almost all of the alternation *Ss* under a 30-sec. intertrial interval. Thus, to minimize position preferences among alternation *Ss* in future studies, it seems advisable to train *Ss* under trials which are as highly massed as possible.

The theory did not fare so well when it was checked against the sequence data reported by Anderson (1960) for nine groups of human *Ss*, each of which received a different sequentially dependent reward schedule. In general, when his R values for Trials 101–300 were ranked according to the qualitative predictions of the theory, the agreement between predicted and obtained ranks was rather poor, particularly for sequences of dissimilar events. For example, when $\pi_{ii} = .70$, Anderson found $R(1) > R(21) > R(121) > R(2121)$ rather than the predicted $R(1) > R(121) > R(2121) > R(21)$. An obvious implication is that the model may require modification if it is to account for human data.

The present findings, coupled with others (e.g., Engler, 1958), testify to the complex effect of sequentially dependent rewards on behavior. Since behavior is apparently a function of more remote events, as well as the immediately preceding event, an analysis of sequence effects is mandatory. Although additional studies with both human and animal *Ss* are required, animal studies seem appropriate at this time. When the *Ss* are animals, *E* can conveniently minimize the effect of more remote events, and thereby simplify his analysis, by merely introducing a series of large intertrial intervals, a procedure that tends to be impractical for human *Ss*.

This advantage should outweigh the inconvenience of position preferences.

Summary

A model derived from statistical learning theory was tested with rats in a T maze. The major independent variables were two sequentially dependent reward schedules and four intertrial intervals. After preliminary training under massed trials, the *Ss* under a given reward schedule were rotated through each intertrial interval in accordance with a latin square design.

During preliminary training, position preferences were minimal for *Ss* under the "repetition" schedule, but not for *Ss* under the "alternation" schedule. Subsequently, only the repetition *Ss* were rotated through the intertrial intervals. The conditional response probabilities for these *Ss* declined as a function of trial spacing. In general, the theory successfully predicted changes in conditional response probability as a function of both trial spacing and various sequences of rewards.

References

Anderson, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

Engler, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

Estes, W. K. Individual behavior in uncertain situations: An interpretation in terms of statistical association theory. In R. A. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954. Pp. 141–149.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

Nicks, D. C. Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

Office of Chief of Ordnance. Tables of the cumulative binomial probabilities. *Ordn. Corps Pamphl.*, 1952, No. ORDP 20–1.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Witte, R. S. A stimulus-trace hypothesis for statistical learning theory. *J. exp. Psychol.*, 1959, **57**, 273–283.

(Received August 4, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 448–455

AN APPLICATION OF STIMULUS SAMPLING THEORY TO SUMMATED GENERALIZATION [1]

TERESA S. CARTERETTE [2]

Indiana University

The general paradigm for experiments demonstrating summated generalization used, for example, by Kalish and Guttman (1957) with pigeons and by Bilodeau, Brown, and Meryman (1956), with human *S*s, is to reinforce a response at two points in a stimulus dimension and, then, to test for generalization along the dimension. The present study represents an attempt to go beyond Bilodeau, Brown, and Meryman by obtaining a more precise determination of the form of the summated gradient for human *S*s and by relating its specific properties to the separation between training stimuli. In addition, the experiment tests a mathematical model for stimulus generalization along a substitutive dimension (Stevens, 1957) derived from the Estes-Burke probabilistic theory of behavior.[3]

Basic to the model is the assumption that, for equal intervals along the dimension, $P(Y) = f(\Delta S)$ is a linear function with negative slope. The dependent variable, $P(Y)$, is the probability of the conditioned response; and (ΔS) is a change along the physical scale of the training stimulus. While Guttman and Kalish's (1956) data for pigeons responding to the wavelength of light generally support the assumption of linearity, human *S*s responding to the position of a light (Brown, Bilodeau, & Baron, 1951) showed such a great drop in response level at the test stimulus immediately adjacent to the training stimulus that clearly most of the gradient was not being adequately sampled. To avoid this high degree of discriminability, the position of the light in the present situation was varied over a continuous field rather than with discrete lamps as in the earlier experiment.

The rationale underlying the simple generalization model and the predictions for summated generalization are elaborated formally in a later section. Briefly, each stimulus in the simple gradient is associated with a set of elements; and some of the elements in each of the sets corresponding to each of the generalized stimuli are assumed to be also present in the set corresponding to the training stimulus. Since the theory relates response probability directly to the proportion of stimulus elements conditioned to a response, the extent to which a test stimulus elicits the reference response reflects the proportion of elements which it shares with the training stimulus. In the summated generalization case where the same response is conditioned to two stimuli in the dimension, response

[1] This paper is based on a PhD dissertation submitted to the Graduate School of Indiana University. The work was executed while the author held a National Institutes of Health Predoctoral Research Fellowship under the sponsorship of W. K. Estes who provided valuable counsel at all stages of the project.

[2] Now at Institute for Psychological Research, Tufts University.

[3] A preliminary presentation of the application of stimulus sampling theory to stimulus generalization phenomena has appeared in mimeographed lectures given by W. K. Estes before the Summer Institute in Social Sciences for College Teachers of Mathematics held at Stanford University in the summer of 1957.

probability should be determined by the proportion of elements which a given test stimulus shares with the two training stimuli.

Application of the assumptions of the set theoretical model (Burke & Estes, 1957) to the simple generalization data obtained in Part I of the experiment provided parameter estimates for the a priori predictions tested in Part II, a summated generalization experiment. Predictions pertained to three aspects of the summated generalization gradients. It was expected that: (*a*) the height at the two training stimuli would be identical to that obtained in the simple generalization experiment, (*b*) midway between training stimuli, the height of the gradients would decrease as the separation between training stimuli increased, and (*c*) the slopes of the outer segments of the summated gradients would be identical for all separation groups.

Method

Subjects

The *S*s were 240 Indiana University undergraduates. Forty-eight provided the data for the simple generalization gradients; and the remainder provided the summated generalization data.

Apparatus

The apparatus enabled *E* to present a 2¾-in. square of light for 1 sec. at different positions along a horizontal line. The *S* faced a black panel which supported a 5 in. high by 44 in. wide field of sand blasted plate glass on which the stimulus light appeared. The walls and ceiling of the room were black; and ambient light was provided by a 7.5-w. frosted ceiling bulb situated directly above *S*.

Design

Simple generalization.—For Group C (Center) the training stimulus was located exactly midway between the left and right limits of the field; and for Group Q (Quarter), it was one-fourth the distance from its left border. The remaining stimuli were to the right of the training stimulus and covered a 3.5-in. range in ½-in. steps.

Summated generalization.—For the Center groups the left-hand training stimulus was at the midway position and for the Quarter groups, at the point one-fourth the distance from the left limit of the stimulus field. The second training stimulus was four ½-in. steps to the right of the first for the T_4 groups; and six and eight steps for the T_6 and T_8 groups, respectively. These separation values were intended to realize two types of situations. In the T_4 groups the stimulus sets associated with the two training stimuli should have a proportion of elements in common. For the T_6 and T_8 groups, these sets should be discrete; but the proportion of elements which the central stimuli share with the training stimuli should differ.

Procedure

Simple generalization.—Each *S* was tested separately in a 40-min. session with 10 min. for dark adaptation and instructions which covered the following points: (*a*) There would be two types of trials: trials on which *E* would announce that the stimulus was to appear in the standard position, and test trials on which the unannounced stimulus might be in the standard or in some different position. (*b*) The stimulus duration would be 1 sec.; and on test trials *S* must decide whether the light had been in the standard or in some different position. (*c*) The two keys in front of *S* were to be used in communicating judgments to *E*. The training stimulus, presented once at the outset of every block of trials, was followed by 8 test trials in which the standard and seven other stimuli each appeared, these blocks continuing to a total of 180 trials. The order of test stimuli within each block was randomized separately for each *S*. To indicate the positive response, half the *S*s were instructed to use the right-hand key and half, the left key. Approximately 9 sec. elapsed between stimulus presentations.

Summated generalization.—Following dark adaptation and instruction, *S* was exposed to 210 trials. Instructions were altered as follows: (*a*) There would be three types of trials: trials on which *E* would announce the stimulus position as *A*, trials on which *E* would announce it as *B*, and test trials on which the unannounced stimulus might be in *A*, or in *B*, or in some different position. (*b*) On test trials if *S* believed that the light was in *A* or *B*, he was to press the right-hand key; and if he believed that it was in neither

position, he was to press the left-hand key. Within each block of 21 trials there were 4 training trials. The two training stimuli each appeared twice, once at the beginning and once midway in the block. The order of test stimuli was randomized with the restriction that each stimulus, including *A* and *B*, appear only once in the 17 test trials.

Results

Simple generalization.—Gradients for Groups C and Q were linear over a 2.5-in. range. Comparison of performance for the first 10 blocks of trials with that for the second 10 blocks did not reveal the inhibitory effects observed by Bilodeau, Brown, and Meryman (1956) who found that as test trials progressed, *S*s produced fewer responses to generalized stimuli. A straight line fitted to the combined data by the method of least squares had a slope of −.16 and its intercept at .82.

Guttman and Kalish (1956) found that gradients for pigeons pecking a key in the presence of different wavelengths of light became flatter as the total number of generalized responses decreased. In the present case three gradients for *S*s grouped on the basis of the total number of positive responses did not have different slopes.

Summated generalization.—Figure 1 presents the mean proportion of positive responses at each test position over the 10 blocks of trials. Response level along the center segments decreases as the distance between training stimuli increases. In the T_4 groups these data points are slightly above those at the training stimuli. For the T_6 groups there is virtually no difference between the training stimuli and the intermediate positions; and in the T_8 groups a decline in response level is clear.

In the absence of normal distributions, nonparametric tests were applied to the data for the entire session. To test the hypothesis that response level at the midway stimulus decreases with increasing separation between training stimuli, each *S*'s mean score at the midway position was subtracted from the mean of his

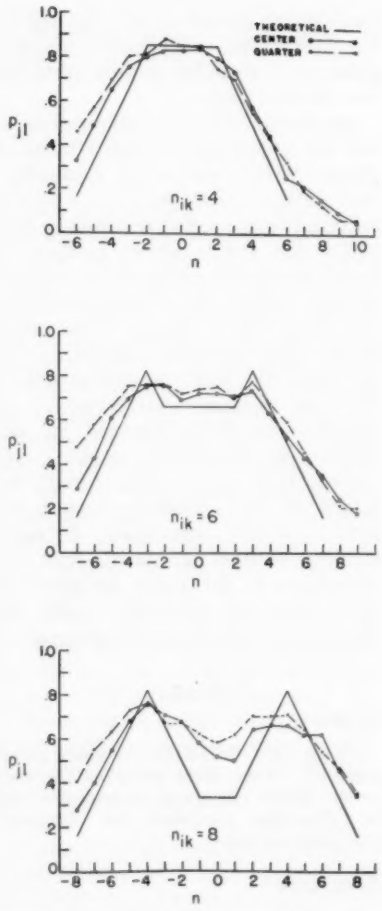

FIG. 1. Theoretical and empirical summated generalization gradients for the T_4, T_6, and T_8 groups. (Zero corresponds to the stimulus exactly midway between training stimuli; and unit changes represent ½-in. steps along the physical scale.)

pooled scores at the training stimuli. The Kruskal-Wallis one-way analysis of variance statistic, H, is significant at the .05 level with 2 *df* for both Group C and Group Q comparisons. The t's comparing the mean proportion of positive responses to the training stimuli with that at the midway stimulus are significant at the .05 level with 62 *df* for the QT_4 and for the T_8 groups only.

The outer segments of all gradients are linear over four stimulus positions. Two measures were used in comparing the slopes of corresponding outer segments: (*a*) each *S*'s mean proportion of positive responses over the four outer stimuli adjacent to a given training stimulus, and (*b*) the difference between each *S*'s pooled mean proportion of positive responses at the first two and second two of these stimuli. The H for Measure 1 on the left side and that for Measure 2 on the right side for Group C gradients are significant at the .05 level with 2 *df*. Although the corresponding outer segments of the gradients for different separation groups are not far from parallel, the possibility of some significant local deviations cannot be ruled out by the present analysis.

Because Group Q gradients are asymmetrical, an effect probably due to the presence of the left border of the frame near this part of the field, tests for symmetry were restricted to Group C data. The Mann-Whitney U is significant at the .05 level for Measure 2 in Group CT_6. The data for Group CT_8 indicate that departures from symmetry in the Center groups may be due to the use of the center position for one of the training stimuli; but any general statement concerning symmetry irrespective of separation between training stimuli is necessarily tenuous.

Although gradients for the first and second halves of the experimental session show no systematic change in response strength for the Center groups, the response level between training stimuli tends to be lower over the second half of the Quarter groups.

Theoretical Analyses

Formal Theory

Correspondence between psychological variables and mathematical sets is defined by the application of the Burke-Estes (1957) component model for stimulus variables. A mathematical set, S_i, is associated with each stimulus. Sets S_0, $S_1, \cdots S_k$ denote $k + 1$ distinguishable stimuli; and the overlap between S_i and S_j represents the similarity between stimuli i and j. The number of elements in any set S_i is N; and the number which S_i and S_j have in common is N_{ij}. Behaviors available to S are represented by A_1 and A_2, the mutually exclusive and exhaustive responses formally distinguished in the experiment; and it is assumed that each element in a set is connected to one of these responses so that, given stimulus i, the probability of A_1 is the proportion of elements in i connected to A_1 and sampled by S on that trial.

The relationship between stimulus and response variables, conceptualized in this manner, has been elaborated in theoretical papers by Burke and Estes (Burke & Estes, 1957; Estes & Burke, 1953). The probability of an A_1 when i is presented is expressed by:

$$p_i = \frac{N(C_i)}{N} \qquad [1]$$

where $N(C_i)$ is the number of elements in S_i connected to A_1.

Generalization of response A_1 to stimulus j, given i as the training stimulus, may be expressed as the sum of the proportion of elements common to S_i and S_j which are conditioned to A_1 by virtue of training on i and the proportion

of elements in S_j only which are connected to A_1 independently of training on i:

$$p_j = \frac{N_{ij}p_i + (N - N_{ij})\rho_j}{N} \quad [2]$$

where ρ_j, a constant for all j, is the probability of an A_1 given j as the stimulus prior to training on i. Substituting w_{ij} for (N_{ij}/N) and rearranging terms, a more convenient form of Equation 2 is obtained:

$$p_j = w_{ij}(p_i - \rho_j) + \rho_j \quad [3]$$

The range over which generalization will occur is determined by the extent of stimulus overlap between the training stimulus and other stimuli in the series. Given training with reinforcement of response A_1 to stimulus i and nonreinforcement elsewhere along the dimension, or in the case of a voluntary response, following instructions to make response A_1 to stimulus i only, it may be assumed that ρ_j equals zero for $j \neq i$. Equation 3 becomes:

$$p_{(i+n)} = w_{i(i+n)}p_i$$

where n is the number of equal intervals on the physical dimension separating i and j; and it may be restated in the form:

$$\begin{aligned} p_{(i+n)} &= p_i - p_i(1 - w_{i(i+n)}) \\ &= p_i - n(1 - w_{i(i+1)})p_i \end{aligned} \quad [4]$$

Equation 4 follows from the fact that stimuli equidistant on the dimension have equal overlaps, so that $w_{i(i+1)} = w_{(i+1)(i+2)} \cdots$ etc. Hence, $n(1 - w_{i(i+1)})$ is the proportion of new elements introduced for a stimulus n steps away from i. If n exceeds $1/(1 - w_{i(i+1)})$, Equation 4 no longer applies.

This equation may be used for predicting generalization along a substitutive dimension. A family of linear gradients may be plotted with p_i as a parameter depending only on learning and n as the independent variable. For experimental convenience it is desirable to determine equal intervals on the physical dimension corresponding to equal decrements in p_i by obtaining a simple generalization gradient empirically. Over the range of physical values for which the gradient fits a straight line, the overlap between i and successive equally separated stimuli may be assumed to decrease by a constant proportion; and the parameters p_i and $w_{i(i+1)}$ may then be estimated from the data and used to test the predictive power of the model in an entirely new situation.

For summated generalization with training on i and k, p_j is the proportion of conditioned elements which j shares with i and k:

$$\begin{aligned} p_j &= \frac{N_{ij}p_i + N_{jk}p_k - N_{ik}p_ip_k}{N} \\ &= w_{ij}p_i + w_{jk}p_k - w_{ik}p_ip_k \\ &= p_i[1 - n_{ij}(1 - a)] \\ &\quad + p_k[1 - n_{jk}(1 - a)] \\ &\quad - p_ip_k[1 - n_{ik}(1 - a)] \end{aligned} \quad [5]$$

For simplicity in notation $w_{i(i+1)}$ has been replaced with a. The quantity $[1 - n_{ik}(1 - a)]$ is the proportion of overlap between the two training stimuli; n_{ik}, the number of steps between i and k; n_{ij} and n_{jk}, the numbers of steps between a test stimulus j, and i and k, respectively. The third term on the right-hand side corrects for the fact that when training stimuli overlap, the sets associated with stimuli between i and k contain elements common to S_i and S_k which have been included twice in the sum of the first two terms.

Two types of relations are determined by n_{ik}. For n_{ik} between zero and $1/(1 - a)$ the training stimuli overlap; and their intersection must be taken into account. If $n_{ik} \geq 1/(1 - a)$, there is no such overlap; and the third term on the right side of Equation 5 is zero. Beyond certain points in the dimension, j will cease to overlap with part of the intersection of i and k, or with either or both training stimuli; and restrictions must be placed on n_{ij} and n_{jk} in both instances. In the present experiment, values of n_{ik} were selected to represent these two types of relations. Given the overlap parameter estimated from the data, training stimuli which are four

steps apart, the value of n_{ik} for the T_4 groups, should overlap. Sets associated with stimuli separated by more than four steps should be discrete; and the T_6 and T_8 groups are intended to represent special cases of this type.

Prediction 1.—For all $n_{ik} \leqslant 1/(1-a)$ the sets associated with test stimuli $j = (i+1)\cdots(i+n-1)$, which lie between i and k, are contained in the union of S_i and S_k; and $p_i = p_j = \cdots p_k$. The solid line in the top panel of Fig. 1 presents this type of gradient, the condition expected to obtain for Groups T_4. It should be noted that when $p_i < 1$, the predicted response strength at the training stimuli following summated generalization training is greater when $n_{ik} < 1/(1-a)$ than when $n_{ik} = 1/(1-a)$. Theoretical gradients for Groups T_6 and T_8 are shown in the other two panels of Fig. 1. Since the training stimuli do not overlap, the central test stimuli should contain elements not in i or k; and generalization at this portion should decrease as n_{ik} increases. The flat region should extend over $(n_{ik}+1)$ points if $n_{ik} < 1/(1-a)$ and over $\left[\frac{2}{1-a} - n_{ik} + 1\right]$ points if $\frac{1}{1-a} \leqslant n_{ik} < \frac{2}{1-a}$.

Prediction 2.—Corresponding outer segments of the gradients should be linear with identical slopes for all n_{ik}. For $n_{ik} < 1/(1-a)$ and $p_i < 1$, the conditioned overlap between i and k produces a departure from linearity which increases as n_{ik} decreases. Gradients should extend over a range of $\left[n_{ik} - 1 + \frac{2}{1-a}\right]$ points.

Adequacy of Predictions for Summated Generalization

The straight line fitted to the simple generalization gradient provided values of p_i and p_j which were substituted in Equation 4 to estimate a. Theoretical and obtained gradients may be compared in Fig. 1.

Prediction 1.—When $n_{ik} = 4$, the discrepancy between theoretical and empirical points along the central segment is in no instance greater than .03. However, the large and consistent disparity between theoretical and obtained values in Groups T_6 and T_8 at these points indicates that predictions from the model are too low when $n_{ik} > 1/(1-a)$.

A salient, and complicating, feature is the downward shift in response strength at the training stimuli. Since, according to the model, response strength at the training stimuli may vary without affecting a, new theoretical gradients might be derived with p_i and p_k estimated from the data for one of the summated generalization groups; but this approach would yield even lower values of p_j than those obtained with $p_i = p_k = .818$.

Statements concerning the symmetry of the center segment and the extent of the plateau must be highly tentative since experimental artifacts discussed in a previous section may have introduced considerable distortion into the gradients. The data do suggest that p_j may be constant over the predicted number of intervals in the central region; but there is no clear indication that the remaining points are symmetrical with those for the outer segments.

Prediction 2.—The outcome of statistical tests comparing the slopes of corresponding outer segments and of tests of symmetry make it impossible to generalize unequivocally concerning the relevance of the results to predictions from the model. On the whole, the data support the assumption that for a substitutive dimension, p_j varies linearly over n. Empirical points over the left outer segments, especially for the Quarter group gradients, generally lie above the predicted values, a disparity which may possibly be due to the conditioning of extraneous cues to the positive response. The ranges over which the empirical gradients extend exceed predictions.

Estimates of a *from summated generalization data.*—The possibility remains that the stimulus component model may be used to extrapolate across summated generalization situations. An additional estimate of a, obtained by fitting a straight line to the data points for the

right hand training stimulus and the six stimuli of the corresponding outer segment of the T_6 groups, yields an overlap parameter equal to .85. In Fig. 2 it can be seen that this estimate results in a better coincidence between theoretical and data points, particularly in the case of Groups T_4 and T_8.

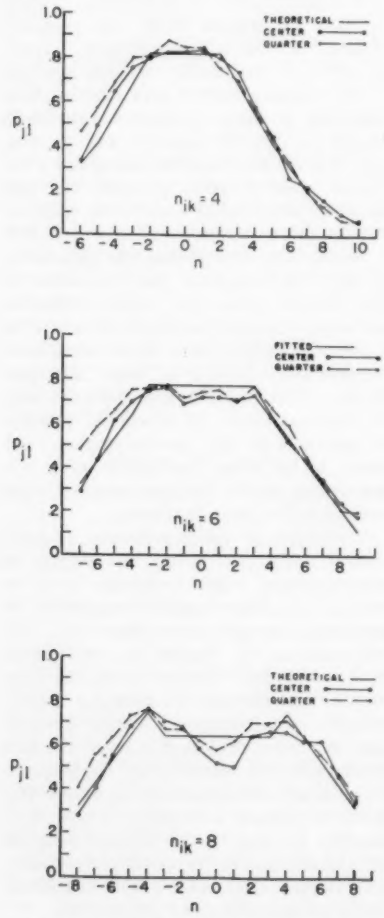

FIG. 2. Theoretical and empirical summated generalization gradients with a estimated from the right hand outer segments of the T_6 groups. (Zero corresponds to the stimulus exactly midway between training stimuli; and unit changes represent ½-in. steps along the physical scale.)

Adequacy of the Theory

In general the two main predictions which follow from the model have been supported. First, the decrease in response level as the separation between training stimuli increases is found in all instances. The second prediction, that corresponding outer slopes of the summated gradients should be parallel for all groups, is well substantiated by the data.

Principal discrepancies between theory and data pertain to three aspects: (*a*) the center dip predicted for the T_6 and T_8 gradients is absent in the T_6 groups and is shallower than expected in the T_8 groups; (*b*) the range of generalization is greater than predicted; and (*c*) the left wings of all gradients, especially in the Quarter groups, are consistently too high. Contrary to prediction and perhaps related to (*a*), the response level for Group QT_4 is significantly higher for the midway stimulus than for tests at the training stimuli. However, it is exactly for the Quarter groups that pronounced asymmetry is present as an experimental artifact. There is the additional possibility that the unequal number of test stimuli on the two outer segments of the T_4 gradients may have interacted with effects of the left border to further increase conditioning to stimuli on the left half of the gradient; consequently, these results remain questionable as legitimate tests of the theory.

If one is mindful of the fact that the parameters used in generating the theoretical curves were estimated from data obtained under different experimental conditions with different *S*s, the model appears to possess considerable predictive power. Certainly the extreme elevation of the left segments of the Quarter groups gradients can be corrected by improvements in the apparatus. The shallowness of the central dip and the underestimation of the range of generalization may be due to the omission of factors which should enter

into the application of stimulus sampling theory to generalization phenomena.

The superiority of predictions based on parameters estimated from a portion of the summated generalization data itself, suggests the possibility that much of the discrepancy between theoretical and empirical gradients in the case of a priori predictions may be attributable to sampling variation associated with the use of two different sets of *S*s in the two experiments.

Summary

Two experiments were designed to test the predictive power of a mathematical model derived from the Burke-Estes (1957) stimulus sampling theory of behavior. The first, an experiment concerned with simple stimulus generalization, yielded estimates of parameters which were used to generate theoretical curves for the second, concerned with summated stimulus generalization. Procedures in the two experiments were almost identical, except that in the summated generalization experiment the training trials occurred in pairs. On training trials *S* was shown a square patch of light at one of the two training positions designated for a particular group. On test trials, if *S* believed the stimulus coincided with either of the training stimuli, he responded "same"; if he believed it differed from both stimuli, he responded "different." Treatments for the six groups differed in two ways: (*a*) there were three degrees of separation between the two training stimuli and (*b*) the range of the stimulus field differed across equal separation groups.

A priori predictions pertained to two aspects of the summated generalization gradients. First, it was expected that increasing separation between the training stimuli would lead to progressively lower response levels along intermediate portions of the gradients. Second, it was predicted that corresponding outer segments of the gradients for different separation groups would be parallel.

The two predicted relationships were generally confirmed. Empirical gradients were approximately linear in their outer segments, varying from clear bimodality in the maximum separation groups to curves with virtually horizontal ledges between the training stimuli in the minimum separation groups. Underestimation of the total range of the summated gradients and the shallowness of the curves between training stimuli constituted the systematic departures from exact predictions. However, when the parameters were estimated from the data for one of the summated generalization groups, theoretical and empirical gradients coincided rather well.

References

Bilodeau, E. A., Brown, J. S., & Meryman, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51**, 293–298.

Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Kalish, H. I., & Guttman, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

(Received August 6, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 456–460

RECOGNITION OF COMPLEX VISUAL STIMULI AS A FUNCTION OF TRAINING WITH ABSTRACTED PATTERNS[1]

MELVIN H. MARX, WILTON W. MURPHY,[2] AND AARON J. BROWNSTEIN[3]

University of Missouri

The present investigation is concerned with the general problem of recognition of complex visual stimuli when such stimuli occur repeatedly in a variety of distorted forms. The problem was orginally suggested by the task faced by the Air Force navigator, who must learn to recognize complex visual stimuli which never look exactly the same on the radar scope, even under optimal conditions. More specifically, we have been concerned with investigating the efficacy of training techniques ("briefing") for recognition of complex visual stimuli in which the pattern, or basic arrangement of the various elements, is the most stable aspect. In such stimuli the peculiarities of shape of individual elements are not stable. The hypothesis tested was that recognition of such stimuli is more effective when *S* is trained on simplified, abstracted versions of the stimulus pattern rather than the variable stimulus itself. The rationale underlying this hypothesis is that elimination of the variably shaped elements eliminates the distraction produced by this factor and thus maximizes *S*'s utilization of the more important pattern cues.

Two methods of abstraction in training were used: (*a*) presentation of a small number of straight lines representing the major direction or "flow" of the pattern, and (*b*) presentation of the pattern represented by simple geometric designs. A control briefing condition consisted of the stimulus form fully drawn in, with variably shaped elements, in the same manner as the test items were prepared. Examples of the training materials are presented in Fig. 1.

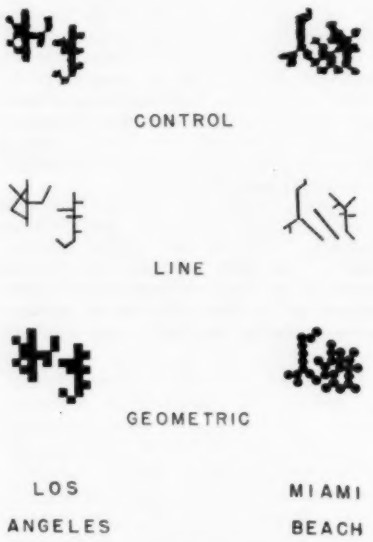

FIG. 1. Samples of training views for two patterns.

[1] This research was supported in part by the United States Air Force under Contract No. AF 41(657)-11, monitored by the Operator Laboratory, Air Force Personnel and Training Research Center, Randolph Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] Now at Wayne State University, Detroit, Michigan.

[3] Now at Florida State University, Tallahassee.

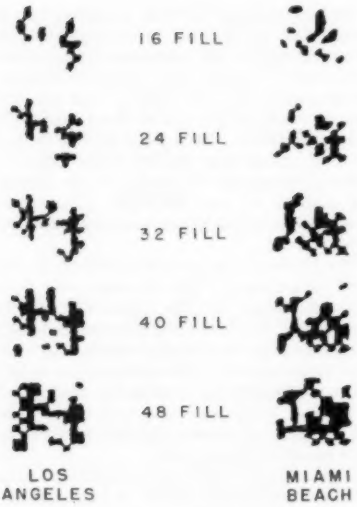

FIG. 2. Samples of test views for two patterns.

Test items consisted of two degrees of perceptual impoverishment (random removal of 25% and 50% of the original stimulus) and two degrees of visual noise (random addition of 25% and 50% irrelevant elements) as well as the complete stimulus itself. Examples of the test views are presented in Fig. 2.

METHOD

Subjects

The *Ss* were 198 men and women students from three general psychology classes at the University of Missouri. Each training condition contained 66 *Ss* and was represented in each class. An additional 65 *Ss* were randomly discarded to equalize group *Ns* to satisfy formal requirements of the analysis of variance. Statistical analyses computed with all *Ss* gave the same results.

Stimulus Materials

Basic patterns.—A set of five basic patterns was prepared in the following manner. A 10 × 10 matrix was used, thus giving 100 available units. Five random and independent selections of 32 units each were made from the 100-unit matrix. The construction rules for these patterns were as follows. The 100-unit matrix was arbitrarily divided into four equal quadrants of 25 units each, the lines of division being the middle, vertical, and horizontal lines. The quadrants were numbered starting with the upper left hand quadrant as number one and proceeding in a clockwise fashion.

One of the four quadrants was picked at random to contain the one focal point, the quadrant diagonal to it being the location of the second focal point. The rules were such that the focal points must be located in these two quadrants. The particular unit in each quadrant was determined by random numbers, the one restriction being that the focal point could not be located in one of the boundary units of the quadrants. The remaining units were then added randomly with the restriction that a new unit must be contiguous with one of the units present but that contiguous boundary units could not be filled.

Control briefing.—Four "artists" were given the five basic patterns, represented by check marks indicating which units of the 10 × 10 matrix were to be filled. They were instructed to fill in the checked units by means of a ball point pen on an Azograph (A. B. Dick duplicating) master sheet, from which several copies of each preparation were made. The only rule was that each filled unit must be drawn at least 50% full, and that no unfilled unit could be drawn more than 50% full. The techniques to be used in connecting units, rounding off corners, etc., were left to the individual artist. Examples of stimuli so prepared are shown at the top of Fig. 1 and in Fig. 2.

Line training.—The same four artists were provided with overlay sheets to place on top of their control drawings. They were instructed to draw in the straight lines that they thought best represented the major parts of the pattern. They were told to keep the lines as simple as possible. Examples of stimuli so prepared are shown in Fig. 1.

Geometric training.—The pattern was depicted by means of geometric symbols, which represented the elements. In any one drawing each unit of the matrix that was to be filled was filled by the same mechanically drawn geometric symbol. Four such symbols were used—dot, empty circle, blacked circle, and blacked square—in order to provide four types of products comparable to the four different artists who prepared the other kinds of briefing stimuli. Examples of such stimuli with blacked squares and circles are shown at the bottom of Fig. 1.

Test views.—Five types of test view were prepared for each pattern. They were: (*a*) the complete 32-unit patterns; (*b*) an incomplete 16-unit view (50% impoverishment); (*c*) an incomplete 24-unit view (25% impoverishment); (*d*) the complete 32-unit pattern plus 8 added units (25% noise); (*e*) the complete 32-unit pattern plus 16 added units (50% noise).

Each of these test conditions was prepared by each artist for each pattern. Independent and random samples were taken from the population of 32 pattern units for each of the 16- and 24-unit views, and from the population of 68 noise units for each of the 8- and 16-unit noise increments. (The population of 68 noise units for any pattern consisted of those units in the matrix that did not comprise any part of the pattern.) Independent random selections were made for each of the artists.

Drawings of the 32-unit pattern itself were made independently of the control 32-unit briefing drawings, so that the 32-unit test and control briefing views were not identical even though their basic pattern was.

Photographs of the 100 test views (5 patterns, 5 types of test view, by 4 artists), as well as the briefing views, were mounted on 2×2 in. slides for projection by an automatic projector.

Experimental Design

Each of the three large classes was divided into three approximately equal sections, with each section receiving one of the three training conditions (control, line, or geometric). The division of a class was done according to columns of seats in an attempt to control for any motivational or ability differences that might be associated with seating arrangements. The 3 *E*s and three classrooms used for the experimental session were rotated over training conditions.

Procedure

The experimental session consisted of 12 training periods each followed by a test trial. In the presentation of the training views each pattern was identified by a city name (New York, Miami Beach, Kansas City, San Francisco, Los Angeles) which was printed plainly beneath the drawing. Test views were to be identified by the *S* by writing in the appropriate initial (NY, MB, KC, SF, LA) on the test sheet.

In each training period each pattern was shown once, in random order. All five slides of any training period were prepared by the same artist, in the case of control and line groups, or utilized the same geometric symbol. Each artist or geometric symbol was represented once in the first block of four training sessions, once in the second block, and once in the third block. Thus, each *S* in the control and line groups was trained three times on materials prepared by each of the artists, and each *S* in the geometric group was trained three times on each of the geometric forms.

The 100 test views were divided into 10 sets of 10 views each. The division was done randomly except for the restriction that in every set of 10 views each pattern and each type of view be represented twice, and each artist at least once. The order of use of test sets was determined randomly, with each set used at least once as a test trial. The first two sets used were repeated in order to make possible 12 test trials.

Each training view was shown for 7 sec., and each test view for 5 sec. A 10-sec. intertrial interval was used.

In the instructions all *S*s were told that they would be shown hypothetical visual representation of five United States cities; that their task was to learn to associate the name of each city with the appropriate pattern so that they would be able to identify it correctly under test conditions; and that after looking at the five city patterns with the names given they would be shown a number of test views, for each of which they were to record on the record sheet the initials of the city they thought was represented.

The *S*s were told that none of the test patterns would be exactly like the training patterns. It was emphasized that some of the test views would resemble the training patterns more than the others, and that some would be missing parts and others would have extra parts.

The line *S*s were told that the artists had made abstracted line drawings from the actual patterns, and geometric *S*s were told that their training patterns would be constructed of simple geometric patterns. The more fully drawn test views were also described to these *S*s.

RESULTS

Figure 3 shows the mean number of correct recognitions for the three training groups as a function of the type of test view, indicated by the number of elements, or amount of

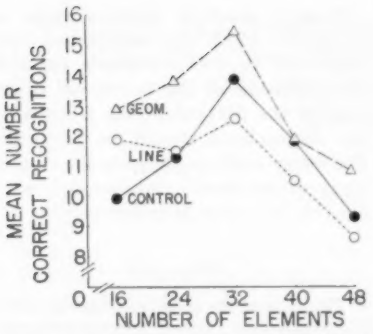

FIG. 3. Mean correct recognitions for the three training conditions as a function of number of elements (amount of fill).

fill, in each view. Table 1 shows the summary of an analysis of variance based on the data presented in Fig. 1. It can be seen that the type of training, number of elements, and the interaction of these variables were all significant sources of variance. The overall means for the three training conditions were: control = 11.29, line = 11.04, and geometric = 13.03. The geometric group mean differed reliably ($P < .01$) from the other two training conditions.

Figure 3 shows that the superiority of geometric training was consistent over the various types of test view. The differences were evaluated by t tests using the error term derived from the major analysis. The geometric group means were superior ($P = .01$) to the other two groups for all types of view except at Fill 16, where the mean differed from that for line briefing at $P < .05$, and at Fill 40, where it failed to differ reliably from the control training.

With regard to the control training-line training comparison, line training was reliably better for the most impoverished test condition ($P < .01$) but control training was better for the 32- and 40-fill conditions ($P < .01$).

The overall trends indicate an asymmetry, with all groups performing better under impoverished test conditions than under noisy test conditions. This asymmetry is especially clear for the abstracted training conditions.

TABLE 1

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT RECOGNITIONS

Source	*df*	*MS*	*F*
Training	2	385.89	9.98*
Error	195	38.65	
Number of elements	4	509.53	87.40*
Interaction	8	40.80	7.00*
Error	789	5.83	

* $P < .01$.

DISCUSSION

The results of this study show that, where *S*s are required to recognize stimuli in which the basic pattern characteristics are more stable than the elements, presentation of certain kinds of abstracted patterns results in greater recognition than presentation of the fully drawn stimuli. This conclusion is especially true for the fragmentary or impoverished type of test where both the geometric and the line abstractions were superior to the control. The superiority in this respect of the severely reduced line stimuli is particularly impressive.

These results are all the more significant in view of the fact that all test items consisted of control-type stimuli. The superiority of recognition in the experimental groups therefore occurred in spite of any positive transfer effects generated by the similarity of control training views and test views. The power of the experimental treatment is suggested by the fact that such an advantage favoring the control condition was so clearly overcome.

From a theoretical point of view, these results are of interest in that they indicate the distracting effect of overloading *S* with cues that are not stable

indicators of the basic stimulus pattern (i.e., the detailed shape of the elements). The extent to which similar abstractions, whereby irrelevant or unstable cues are eliminated, may be effective in other types of learning needs to be investigated. It is important to remember that the generality of the present results is restricted to situations in which the eliminated cues are of low or questionable usefulness. The general suggestion, however, is that wherever such questionable cues can be reduced or removed acquisition efficiency may be facilitated. In further investigations the degree of stability of such cues should be directly manipulated.

Another approach to the problem is that the geometric presentation during training stood in a prototypic role to the subsequent test material in the matrix context. That is, this type of training clearly identified the matrix cells involved in the stimulus pattern. Recognition behavior would be expected more readily to generalize to variations of the prototype following learning of the prototype than following learning of similar variations of the prototype, as was the case in control training.

From a practical point of view, the extent to which the abstraction technique will be useful depends upon the determination of the functional dimensions of the particular task along which the abstractions should be performed. The above analysis suggests the employment of prototypic information where possible, as a first approximation.

Summary

The *S*s were shown different presentations of complex visual stimuli and tested on their ability to identify distorted versions of the stimuli. A simple geometric technique, in which the fundamental stimulus pattern was uniformly represented by dots, circles, or squares, was found to be consistently superior as a means of training. Training which utilized only the main lines of the patterns was found to be more effective than the control (fully drawn) training stimuli for certain test items of a fragmentary sort. For test items in which the trained pattern was embedded in noise, control training was relatively more effective than line, although even here the geometric training stimuli were clearly superior.

(Received August 11, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 461–467

PARTIAL REINFORCEMENT EFFECTS (PREs) AND INVERSE PREs DETERMINED BY POSITION OF A NONREWARDED BLOCK OF RESPONSES[1]

GLEN D. JENSEN

University of New Mexico

In a runway experiment (Jensen & Cotton, 1960), a partial reinforcement effect (PRE) during Day 10 was observed in a group of rats which received 30 rewarded (R) trials on Day 1, 15 nonrewarded (NR) trials followed immediately by 15 R trials each day during Days 2–9, and 30 NR trials on Day 10. This group will be referred to as 15NR-15R. Another group (30NR-30R) which received 30 R trials on Days 1, 3, 5, 7, and 9 but 30 NR trials on Days 2, 4, 6, 8, and 10 did not show a PRE during Day 10. In the case of both of these groups, the PRE was determined in relation to a group (Control) which received 30 R trials during Days 1–9 and 30 NR trials on Day 10.

These results are consistent with the interpretation that NR trials must be patterned in such a way with R trials as to lead to the hypothesis or expectancy that reward comes after a period of nonreward and during the same session as the nonreward, if the NR trials are to produce a PRE. The 15NR-15R *S*s had eight opportunities to form such a hypothesis or expectancy prior to Day 10. The fact that these *S*s ran progressively faster during the last five NR trials of a session as Day 10 approached is consistent with the notion that they were gradually forming the expectancy that reward comes after the initial NR trials of a day are completed. Actually, another assumption is implicit: namely, that a heightened motivational state goes with this hypothesis or expectancy.

Grosslight and his associates (Grosslight & Child, 1947; Grosslight, Hall, & Murnin, 1953; Grosslight & Radlow, 1956, 1957) have marshaled considerable evidence that blocks of NR trials followed by blocks of R trials produce a PRE. However, they used very small blocks of NR trials, and their presentation does not show the gradually increasing vigor of responding during successive blocks of NR trials which seems to characterize this schedule.

Experiment I

Three of the groups of Exp. I constitute essentially a replication of the three runway groups already discussed, except that a Skinner box was used and hence single bar presses were substituted for single runway trials. The fourth group, however, underwent conditions which should produce the *inverse* of the traditional PRE if the expectancy interpretation outlined above is correct. The rats received a schedule of 20R-20NR for several days, which should result in slower pressing during the last (NR) session relative to 100% control *S*s, if it is assumed that a depressed motivational state goes with the expectancy of nonreward.

Grosslight et al. ran several groups comparable to Group 20R-20NR of

[1] This research was supported in full by Grant M-2834 from the United States Public Health Service.

the present study, except that they used very small blocks of NR trials. In no case did they obtain a significant inverse PRE, but one paper (Grosslight & Radlow, 1956) presents an effect which is suggestive of an inverse PRE. It was hoped that the large blocks of NR trials presently employed would produce a significant inverse PRE and also that gradually decreasing vigor of responding would occur during the NR presses on successive days of 20R-20NR scheduling.

Method

Subjects and apparatus.—The *S*s were 40 male albino rats of the Sprague-Dawley strain from 68 to 134 days in age. All *S*s were Skinner-box naive, but most of them had received 1–3 hr. of training in a runway.

A Foringer small animal test chamber was used along with Foringer programing equipment, a Gerbrands cumulative recorder, and a Davis 45-mgm. pellet dispenser.

The Gerbrands recorder was operated at the standard paper speed of 11 in. per hr. An event pen was wired through a Foringer-built contingency selector panel in such a way that every tenth bar press caused the event pen to make a horizontal mark on the paper. The distance between successive marks was used to determine the total amount of time expended during a block of 10 presses.

Procedure.—For 1 to 4 weeks before the Pretraining Day all *S*s were fed once a day with a ration of 10 gm. of finely-ground Purina laboratory chow. During the last 2 days before the Pretraining Day, all *S*s got their ration 1 hr. after the hour of the day at which they were to receive pretraining and training. The *S*s had access to water at all times except during pretraining and training sessions.

On the Pretraining Day each *S* was placed in the box for 55 min. with the apparatus set to deliver one pellet every 30 sec. until a total of 110 pellets were received. Immediately after the 55-min. session *S* was fed the remainder (5 gm.) of his usual 10-gm. ration in the form of ground chow.

Twenty-four hours after the beginning of the pretraining session each *S* was placed in the box with the bar connected to the pellet dispenser and with the automatic-delivery circuit broken. Each *S* was allowed to make 20 R presses and was then returned to his home cage where he received 9 gm. of ground chow 1 hr. after the beginning of the session. This session will be called Day 1.

The 40 rats which successfully completed Day 1 were randomly assigned to four experimental conditions except that the last 3 *S*s in a group were placed so that the four groups would have approximately equal means and variances with respect to the second 10 presses of Day 1. Group R received 40 R presses on each of Days 2–5. Group 20NR-20R received 20 NR presses followed immediately by 20 R presses during each of Days 2–5. Group 40NR-40R received 40 NR presses on Days 2 and 4, and 40 R presses on Days 3 and 5. Group 20R-20NR received reward during the first 20 presses of a day during Days 2–5 and no rewards during the last 20 presses of each of these days. All groups received a 55-min. session with no press rewarded on Day 6. All groups were 22 hr. food deprived at the beginning of each session and the 10-gm. daily ration was maintained throughout the 6 experimental days.

The time to the nearest .01 sec. taken by each *S* to make each block of 10 successive presses was determined and these values were transformed to their reciprocals divided by 10. These transformed values will henceforth be called "rate" scores.

Results and Discussion

Figure 1 shows the mean rate data for each group through the entire course of experimentation.

The representation of the long NR session given during Day 6 is complicated by the fact that *S*s in the same group made widely different numbers of presses. Arbitrarily, it was decided to represent each group for as many presses as the number made by the *S* who made the most presses. If, for example, the longest pressing *S* in Group 20NR-20R made 500 presses, and if another *S* made 200 presses, the latter *S* was assigned the arbitrary minimum rate value of .0083 for each block between 200 and 500. Note that five rate scores for each *S* in a group have been averaged in the determination of each point plotted for this 55-min. NR session. Since each *S* made at least 100 presses, the first two points are not arbitrary.

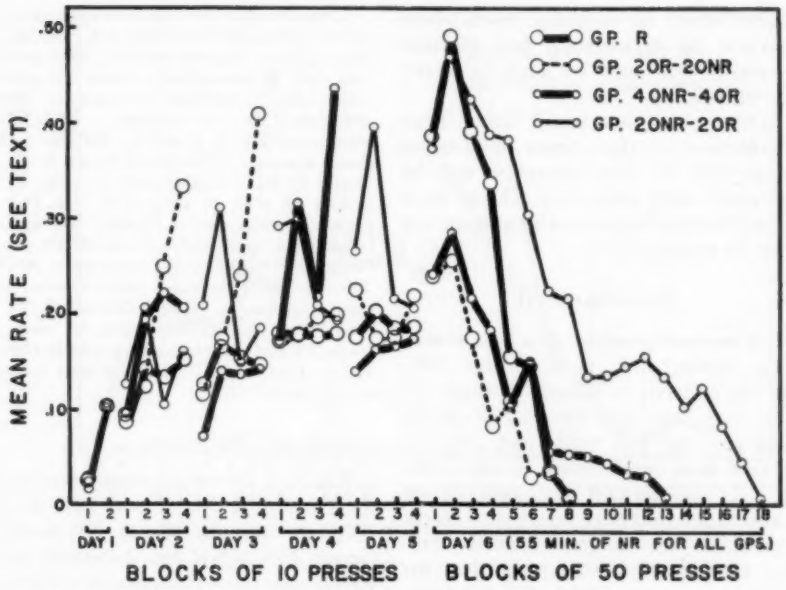
GP. R
GP. 20R-20NR
GP. 40NR-40R
GP. 20NR-20R
MEAN RATE (SEE TEXT)
.50
.40
.30
.20
.10
0
1 2 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
DAY 1
DAY 2
DAY 3
DAY 4
DAY 5
DAY 6 (55 MIN. OF NR FOR ALL GPS.)
BLOCKS OF 10 PRESSES
BLOCKS OF 50 PRESSES

FIG. 1. The course of mean rate of pressing during all phases of experimentation.

Later data points are more or less representative of the number of "persistent" *S*s in a group as well as their rates of pressing.

The least arbitrary, but not necessarily the most sensitive test for a PRE and for an inverse PRE involves the use of the total number of presses made during the 55-min. session on Day 6. Analysis of variance on these scores yielded an F of 4.71 ($P < .05$; 3/36 *df*). A subsequent Tukey[2] analysis to compare individual means revealed that the 20NR-20R animals made significantly more presses than each of the other groups at the .05 level or better, but no other comparisons were significant. Thus, by this measure there is a PRE in the 20NR-20R group but there is not the predicted inverse PRE in Group 20R-20NR. There is, however, a fairly large difference in mean numbers of presses made between Groups R and 20R-20NR, namely 220.8 and 159.8, respectively.

Figure 1 shows that Groups 20R-20NR and 40NR-40R pressed at a markedly slower rate early in Day 6 relative to the other groups. An analysis of variance using the mean rate for each *S* during presses 0–100 (as determined by averaging the first 10 rate scores for each *S*) revealed that the groups differed significantly at the .03 level. A Tukey analysis revealed that Groups R and 20NR-20R showed a higher rate at the .04 level or better relative to the other two groups and no other comparison were significant. Thus both Group 20R-20NR and Group 40NR-40R showed an inverse PRE on this response measure but Group 20NR-20R did not show a PRE. If the first 300 presses were used, Group 20NR-20R probably would show a PRE, but

[2] J. W. Tukey. The problem of multiple comparisons. Privately circulated monograph, 1953.

there would be so many cases where several *Ss* in a group had stopped pressing that such an analysis would be misleading.

Other analyses could have been performed on data from days other than Day 6, but comment will be deferred until after Exp. II has been described and more reliable statements can be made.

Experiment II

It seemed possible that the slower rate during Day 6 of Group 20R-20NR relative to Group R might be due to the fact that Group 20R-20NR had received NR trials just prior to Day 6 and not because Group 20R-20NR had formed the concept or expectancy that once nonreward starts there is no further reward during that session. Consequently, two groups were run which differed from Groups R and 20R-20NR of Exp. I only in that both groups made 20 R presses immediately prior to a 55-min. NR session. Again it was predicted that the new 20R-20NR condition would produce an inverse PRE.

Three additional groups were run to determine whether a more pronounced PRE could be produced by giving seven 20NR-20R experiences rather than the four experiences given in Exp. I and to determine whether a more pronounced inverse PRE could be produced by giving six 20R-20NR experiences.

Method

Subjects and apparatus.—The *Ss* were 50 male albino rats of the Sprague-Dawley strain from 69 to 131 days of age on the pretraining day. The *Ss* had all received 1–9 hr. of experience in a runway and hence were quite docile.

Procedure.—The procedure was identical to Exp. I through the 20 R presses of Day 1. The first 6 *Ss* of a group were assigned randomly, but the remaining 4 *Ss* were placed so that the five groups of Exp. II would be comparable to each other as well as to the four groups used during Exp. I.

Group R No. 2 was allowed to make 40 R presses during each of Days 2–4. On Day 5 these *Ss* were allowed to make 20 R presses and then *E* immediately threw the switch which initiated a 55-min. NR session. Group 20R-20NR No. 2 received a 20R-20NR session on Days 2, 3, and 4. On Day 5 this group was treated exactly as Group R No. 2. Group R No. 3 was allowed to make 40 R presses on each of Days 2–7. On Day 8 these *Ss* underwent a 55-min. NR session immediately after making their 20th R press. Group 20NR-20R No. 2 received a 20NR-20R session on Days 2–8 and a 55-min. NR session on Day 9. Group 20R-20NR No. 3 received a 20R-20NR session on each of Days 2–7 and were treated exactly as Group R No. 3 on Day 8. The data were treated exactly as in Exp. I.

Results and Discussion

Figure 2 shows the mean rate data for Groups R No. 2 and 20R-20NR No. 2 through the 55-min. NR session. Group 20R-20NR No. 2 showed rates which were quite comparable to Group 20R-20NR of Exp. I during every phase of experimentation. Group R No. 2, however, showed markedly slower pressing during the 55-min. NR session relative to the rate exhibited by Group R of Exp. I. Analysis of variance using the total number of presses made during the 55-min. NR session yielded an F of less than 1.0. Thus no significant inverse PRE was demonstrated by this response measure.

Another analysis of variance was done using the mean of the first eight 10-press rates during the 55-min. NR session. One such score was obtained for each of the 20 *Ss* in Groups R No. 2 and 20R-20NR No. 2. These scores represent the largest number of presses which can be used without using arbitrary scores for some *Ss*. The F obtained is only .49, which is below the mean of the F distribution. Thus no inverse PRE was demonstrated by this measure either. However, the fact that Group 20R-20NR No. 2 showed rates very comparable to those of Group 20R-

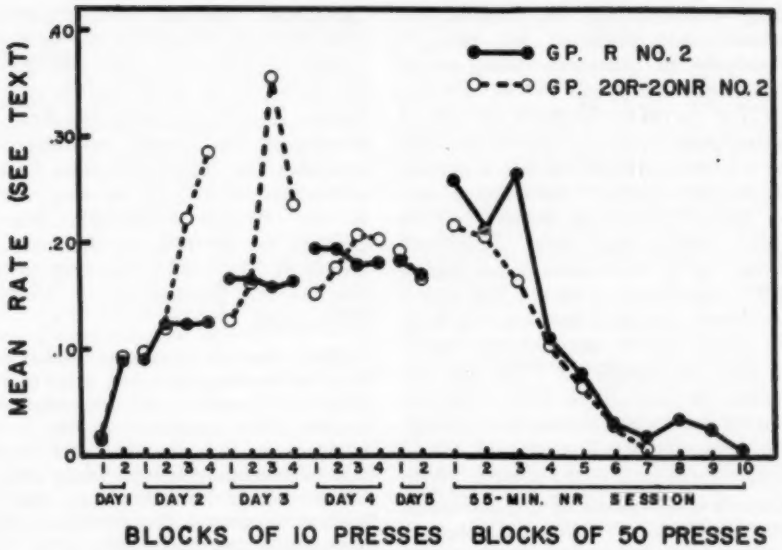

FIG. 2. Mean rate during all phases of experimentation for Groups R No. 2 and 20R-20NR No. 2.

20NR of Exp. I is encouraging, and the failure to get an inverse PRE in this part of Exp. II may yet prove to be due to the fact that Group R No. 2 was spuriously slow during nonreward.

Figure 3 shows the mean rate data for the other three groups of Exp. II. Note the close correspondence between the rates of Group 20R-20NR No. 3 and Group R No. 3 on all R blocks, especially the second R block of each day. An analysis of variance using the total number of presses made during the 55-min. NR session for each *S* yielded an *F* of 5.40 ($P < .015$; 2/27 *df*) between the three groups. A subsequent Tukey analysis revealed that Group 20NR-20R No. 2 made more presses than either of the other groups at the .02 level or better, and no other comparisons were significant. Thus, as in Exp. I, the 20NR-20R condition produced a significant PRE but the 20R-20NR condition did not produce a significant inverse PRE when total number of NR presses during 55 min. serves as the criterion. The *S*s in Group R No. 3 made an average of 240.3 presses whereas *S*s in Group 20R-20NR No. 3 made 204.4 presses on average. The percentage difference is smaller than in Exp. I, so it must be concluded that the two additional 20R-20NR experiences did not increase the inverse PRE with respect to this response measure. The *S*s in Group 20NR-20R No. 2 made an average of 375.4 presses. The percentage difference relative to Group R No. 3 is smaller than the corresponding figure from Exp. I, so it must be concluded that the extra three 20NR-20R experiences did not produce a larger PRE. These findings may mean that some value less than or equal to four 20NR-20R or 20R-20NR experiences is necessary and sufficient to produce either the PRE or the inverse PRE at the full strength, when total number of presses is the criterion.

The mean of the first eight 10-press

rate scores during the 55-min. NR session was obtained for each *S*. Analysis of variance using these scores resulted in an F of 3.41 ($P < .05$; 2/27 *df*) between the three groups. A subsequent Tukey analysis revealed that Group 20R-20NR No. 3 pressed slower than each of the other groups at the .03 level or better, but no other comparison was significant. Thus, using this measure, an inverse PRE was obtained in the 20R-20NR condition just as it had been in Exp. I. These results also confirm Exp. I in that no significant PRE was obtained in the 20NR-20R condition during these early presses even though the difference between relevant groups is very large later in the session. With respect to the effect of the additional 20R-20NR and 20NR-20R experiences, the differences are at least in the right direction. The 20R-20NR *S*s of Exp. I pressed at 56.97% of the rate of the R *S*s during the first 100 presses of Day 6: Group 20R-20NR No. 3 pressed at 55.56% of the rate of Group R No. 3 during the first 80 presses of the 55-min. nonrewarded session. The 20NR-20R *S*s of Exp. I pressed at 96.96% of the rate of the R *S*s: Group 20NR-20R No. 2 pressed at 119.99% of the rate of Group R No. 3 with reference to the first 80 NR presses of the 55-min. NR session.

There are still some phenomena evident in the data gathered prior to the 55-min. NR session which warrant discussion. Note, especially in Fig. 3, the tendency for *S*s in the 20NR-20R condition to press progressively faster during the second 10 NR presses as days of training progress. No statistical tests were made, but the graphs show that the actual trend is very strong and very consistent in the two 20NR-20R groups.

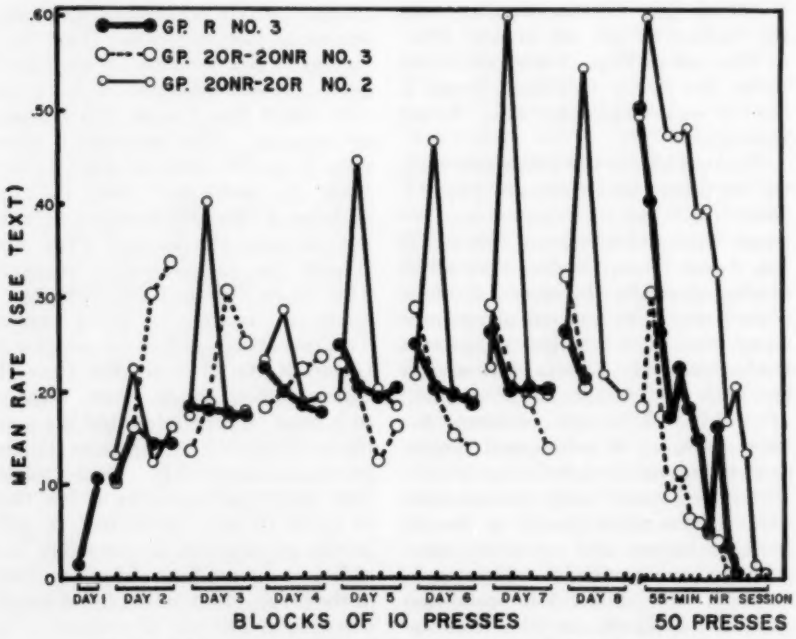

FIG. 3. Mean rate during all phases of experimentation for Groups R No. 3, 20R-20NR No. 3, and 20NR-20R No. 2.

This phenomenon is in every way comparable to the trend noted in the 15NR-15R group in the runway study by Jensen and Cotton (1960). Both findings are consistent with the notion that the rat makes hypotheses based upon past experience with respect to schedules of reinforcement. At least this is so if we further assume that the hypothesis caused by a __NR-__R schedule is one which results in a heightened motivational state. On the other hand, the 20R-20NR conditions showed a strong day-by-day trend toward *slower* pressing during both the first and second block of 10 NR presses of a day as training progresses. This phenomenon has been demonstrated three times and hence statistical tests would seem to be superfluous.

The fact that the same number and proportion of NR responses can produce either a PRE or an inverse PRE would appear to have important implications bearing upon the understanding and prediction of behavior under all conditions of partial reinforcement. If it is true that animals characteristically make hypotheses about schedules of reinforcement, as they appeared to do in the present experiments, then psychologists will not fully understand partial reinforcement until they find the general laws whereby these hypotheses are formed as well as the laws relating these hypotheses to vigor and persistence of responding. This problem can be reduced operationally to the parametric study of a number of patterning variables such as the number of R and NR presses in daily __R-__NR and __NR-__R schedules and also the number of switches from R to NR blocks of presses within sessions involving more than one switch.

Summary

Nine groups of 10 rats each were run through conditions of successive blocks of rewarded (R) and nonrewarded (NR) bar presses in a Skinner box. All *Ss* received 20 rewarded presses on Day 1 and the reciprocals of the time spent making the last 10 presses were used to match groups. In one condition *Ss* received a pattern of 20NR-20R during each of Days 2–5 and a 55-min. NR session on Day 6. Another group received this schedule during Days 2–8 and then a 55-min. NR session during Day 9. Both groups showed a strong day-by-day trend to press faster during the last 10 NR presses of a day, and both displayed very high rates of pressing during the 55-min. NR session relative to control groups which received 40 R presses for as many days as these groups received 20NR-20R scheduling. These results are consistent with the notion that the 20NR-20R *Ss* gradually formed the hypothesis that reward follows nonreward and that a heightened motivational state goes with the hypothesis.

One group received a schedule of 20R-20NR during each of Days 2–5 and a 55-min. NR session on Day 6. Another group got the same treatment except that the 55-min. NR session came immediately after 20 R presses on Day 5. A third group received the 20R-20NR schedule during Days 2–7 and a 55-min. NR session immediately after 20 R presses on Day 8. All three 20R-20NR groups showed a day-by-day trend toward slower pressing during the 20 NR presses of a day and all showed slower rates of pressing during the first 80 or 100 NR presses of the 55-min. NR session relative to the appropriate 100% control groups. These results are consistent with the interpretation that these *Ss* gradually formed the hypothesis that reward does not follow nonreward within the same day and that a depressed motivational state goes with the hypothesis.

References

Grosslight, J. H., & Child, I. L. Persistence as a function of previous experience of failure followed by success. *Amer. J. Psychol.*, 1947, **60**, 378–387.

Grosslight, J. H., Hall, J. F., & Murnin, J. Patterning effect in partial reinforcement. *J. exp. Psychol.*, 1953, **46**, 103–106.

Grosslight, J. H., & Radlow, R. Patterning effect of the nonreinforcement-reinforcement sequence in a discrimination situation. *J. comp. physiol. Psychol.*, 1956, **49**, 542–546.

Grosslight, J. H., & Radlow, R. Patterning effect of the nonreinforcement-reinforcement sequence involving a single nonreinforced trial. *J. comp. physiol. Psychol.*, 1957, **50**, 23–25.

Jensen, G. D., & Cotton, J. W. Successive acquisitions and extinctions as related to percentage of reinforcement. *J. exp. Psychol.*, 1960, **60**, 41–49.

(Received August 15, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 468–475

THE RELATIONSHIP BETWEEN KINESTHETIC SATIATION AND INHIBITION IN ROTARY PURSUIT PERFORMANCE

RONALD S. LIPMAN [1] AND HERMAN H. SPITZ [2]

Edward R. Johnstone Training and Research Center

Several writers (Duncan, 1956; Eysenck, 1955; Köhler & Fishback, 1950) have suggested a similarity, if not an identity, between the process of satiation postulated by Köhler and Wallach (1944) to explain certain displacement effects in perception, and the process underlying some form of inhibition in learning.

Köhler and Wallach (1944) hypothesize that the act of fixating a "figure" sets certain electrochemical processes in motion in the brain, isomorphic to the contours of the figure, which, over time, create an area of impedance (satiation) that spreads in diminishing strength through the immediately surrounding brain field. As a consequence of satiation, subsequent figural currents are blocked and displaced resulting in various types of perceptual "distortion." Köhler and Wallach have employed this theory of brain processes to explain the occurrence of figural aftereffects, reversible figures, and certain illusions.

[1] The authors would like to express their grateful appreciation to Marshall P. Smith, Chairman, Psychology Department, Trenton State College, for his aid in enlisting volunteers and making space available for Exp. II. We are grateful also to Leonard S. Blackman, Belver C. Griffith, and Philip J. Siegmann of the Edward R. Johnstone Training and Research Center, Rupert Hester of the Bureau of Social Research, Department of Institutions and Agencies, State of New Jersey, and David Zeaman and Betty House of the University of Connecticut, all of whom assisted in some aspects of the present study.

[2] Now at Special Studies Unit, National Institute of Mental Health, Bethesda, Md.

In Hullian (Hull, 1943) theory, inhibitory potential ($\dot{I}_R$) consists of two additive components: reactive inhibition (I_R) and conditioned inhibition ($_sI_R$). These constructs, extended by Kimble (1949a) and others to motor learning, have been employed to explain such seemingly diverse phenomena as reminiscence, spontaneous recovery, alternation, extinction, disinhibition, and the superiority of spaced over massed motor practice. I_R is likened to a primary negative drive state that develops whenever a motor response is made, tends to block the recurrence of the response and to dissipate spontaneously with rest. $_sI_R$, on the other hand, can be viewed as a "resting" habit in which stimuli present during I_R reduction become conditioned to I_R and when present in the learning situation tend to weaken the excitatory potential of stimulus situations otherwise evoking positive responses.

Among the writers who have suggested a similarity between satiation and inhibition, Köhler and Fishback (1950) did not identify their reference to inhibition with a particular learning theory, whereas both Duncan (1956) and Eysenck (1955) specifically referred to Hull (1943). Duncan (1956), however, rejects Hull's $_sI_R$ as a necessary construct, and limits his discussion to the similarity between Hull's I_R construct and satiation, both processes being viewed as central in locus. Eysenck (1955), for unspecified reasons, also limits his anal-

ogy to I_R. He postulates that the high satiater is the individual ". . . in whom reactive inhibition is generated quickly, in whom strong reactive inhibitions are generated, and in whom reactive inhibition is dissipated slowly . . ." (p. 97). The low satiater develops I_R slowly, in lesser amounts and dissipates it more quickly.

The hypothesis of a similarity between satiation and inhibition has been investigated by Meier (1956) and Rechtschaffen (1958). For reasons to be discussed more fully in a later section, the results of these studies were not in agreement.

The present study reports the results of two experiments in which the rotary pursuit performance of high and low kinesthetic satiaters was compared under different work and rest schedules in an attempt to more fully test the similarity hypothesis and to identify the Hullian analogue of satiation in motor learning.

General Method

Subjects.—Retarded adolescents were used in Exp. I. College sophomores were employed in Exp. II.

Task order.—In both experiments *S*s were first given the kinesthetic figural aftereffect test (KAE). The *S*s with extreme KAE scores were then tested on the rotary pursuit. Each experiment was completed within a 4-mo. period.

KAE.—Since Köhler and Dinnerstein's (1947) original suggestion that kinesthetic aftereffects may be due to a satiation process, KAE has been employed widely as an index of kinesthetic satiation. Briefly, *S* is required to make equality of width judgments with the same block before and after rubbing a larger (or smaller) block. For example, when *S* rubs the larger block, the standard block feels narrower. The amount of displacement is taken as an index of the strength of the satiation process.

The KAE apparatus employed in the present study was similar to the one pictured by Klein and Krech (1952). It was mounted on two wood horses set approximately 24 in. apart. Depending on *S*'s height, *S* either stood, or sat on a stool, between the horses. A variable test (T) block, on which the position of a rider could be read to the nearest $\frac{1}{16}$ in., was situated on *S*'s right. The rider was free to move along the block from the narrow end of $\frac{3}{4}$ in. to a wide end of $2\frac{1}{4}$ in. The block was constructed so that a 1-in. change in length corresponded to a $\frac{1}{16}$-in. change in width. On *S*'s left, a rectangular T block $1\frac{1}{2}$ in. wide could be quickly replaced by a rectangular inspection (I) block $2\frac{1}{4}$ in. wide. All blocks were made of smooth hard wood, were $\frac{13}{16}$ in. thick, and had riders. The *S*s were shown the apparatus and told that they would be required to make equality of width judgments. It was demonstrated that one T block was of fixed width whereas the other T block was thin at its lower end and gradually widened. The *S* was told that he would be asked to hold the rectangular T block between the forefinger and thumb of his left hand, and the variable T block in the same manner with his right hand. By moving the right-hand rider either up or down the variable T block, he was to "find the spot on the right hand bar that feels as wide as the left hand bar." The college *S*s were then given two practice trials, while the retarded *S*s were given four practice trials (to reduce their greater variability), with feedback to the point of objective equality (POE). All trials on the KAE were alternated from random positions below and above POE. Position habits were minimized by randomly shifting the position of the variable T block before each judgment.

The normals were blindfolded before the initial practice judgments, the retardates after (for reasons of rapport). Then all *S*s made four additional judgments without feedback. The algebraic mean of these settings defined *S*'s point of subjective equality (PSE). Next, *S* firmly grasped the inspection bar and rubbed back and forth for $1\frac{1}{2}$ min. in time to a metronome set at 108 beats per min. The test bar was then quickly replaced and four more equality judgments made. The postinspection PSE was subtracted from *S*'s pre-inspection PSE to yield *S*'s mean displacement, or satiation score. These scores were measured in $\frac{1}{16}$-in. units; a + sign indicated that *S*'s postinspection PSE was smaller (in the expected direction) than his pre-inspection PSE, a − sign, the reverse.

Rotary pursuit.—A standard 60 rpm pursuit rotor and flexible stylus, with a silver head $\frac{1}{8}$ in. in diameter, were employed. Time on target scores were recorded in .001 min.

by a standard 6v. dc timer. The brass target was $\frac{3}{4}$ in. in diameter and set flush $2\frac{1}{4}$ in. from the edge of the black disc, which was $11\frac{13}{16}$ in. in diameter. The rate of rotation was checked daily and the stylus head and target were polished before testing each *S*. The *S*s were given standard instructions with *E* demonstrating how to hold the stylus and the required circular motion.

Experiment I

Method

Subjects and procedure.—Two KAE scores, separated by an interval of not more than 3 weeks, were obtained on 112 high-grade retardates (IQ above 60). The test-retest Pearson *r* was low (+.34) but significant ($P < .01$). The *S*s whose KAE scores were consistantly high or low on both administrations were selected for the rotary pursuit task. Group HS (high satiation) consisted of 23 males whose mean displacement scores were 3.75 mm. and 3.89 mm. Group LS (low satiation), 24 males, had mean displacement scores of −.48 mm. and −.03 mm. There was no overlap of satiation scores among the two groups. A third group—Group S (spaced)—of 18 males was selected to provide a close IQ and age match to Groups HS and LS. The mean IQs of Groups HS, LS, and S were 73, 71, and 72, respectively; the mean age, 16–6, 17–7, and 17–8.

Rotary pursuit schedule.—The massed to spaced performance schedule for Groups HS and LS consisted of 20-sec. work and 10-sec. rest for 33 trials, a 60-sec. rest, followed by 15 additional 20-sec. trials with 60-sec. intertrial rest. Group S received 48 trials of 20-sec. duration with 40-sec. intertrial rest.

Results

The results are displayed in Fig. 1. Over Trials 12–33, Group S, as anticipated, performed at a higher level than both Group LS ($P = .02$) and Group HS ($P < .001$). The Mann-Whitney *U* test was employed for all comparisons unless otherwise stated and the reported *P* values are all based on two-tailed tests. Group LS over this same block of trials performed at a significantly higher level than Group HS ($P = .02$).

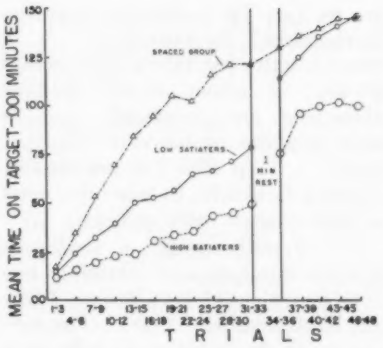

FIG. 1. Mean time on target scores for high and low satiaters under a massed to spaced schedule, and for a control group under a spaced schedule. (Trials are plotted in blocks of three. One-minute rest for massed to spaced groups only.)

The reminiscence gain (the difference between the first three spaced trials and the last three massed trials) was also greater in Group LS than in Group HS ($.10 > P > .05$). Finally, a comparison of the three groups over Trials 46–48 revealed that Groups LS and S were on target significantly more often than Group HS ($P = .02$ for both comparisons).

These results were somewhat complicated by the fact that Group LS performed at a higher level ($.10 > P > .05$) on the initial trial. In order to control for starting differences, an analysis of covariance was performed for Trials 12–33 and Trials 46–48. The corresponding *F*s of 5.27 and 7.59 (for *df* 1/44) were associated with *P* values below the .05 and .01 levels.

Discussion

These data support the similarity hypothesis. The high satiater develops more inhibition than the low satiater under the massed schedule and tends to

dissipate less inhibition than the low satiater when switched to the spaced schedule.

The lower performance of Group HS over Trials 12–33 is difficult to interpret since it may have been attributable to a faster rate of accumulation of inhibition during work and/or a slower rate of dissipation of inhibition during the 10-sec. intertrial rest. If all inhibition is viewed as I_R, in keeping with Duncan's (1956) formulation, the reminiscence difference suggests that the dissipation of I_R during rest is a crucial factor. To test whether high satiaters also build in more I_R during work, as postulated by Eysenck, a pure massed schedule is needed.

Within Hull's two-factor inhibition theory, the most impressive finding of the present study is the asymptotic difference of Group HS relative to Group LS and to Group S. The low satiaters have not developed ${}_sI_R$ whereas the high satiaters are characterized by a significant amount of ${}_sI_R$. This, in turn, indicates that the I_R developed during massing was composed of only I_R in the low satiaters, whereas it was composed of both I_R and ${}_sI_R$ in the high satiaters. The presence of ${}_sI_R$ in Group HS but not in Group LS may have resulted from one or a combination of the following factors: (*a*) the rate at which I_R develops during work; (*b*) the level of I_R tolerated before conversion to ${}_sI_R$ takes place; or (*c*) the rate at which I_R dissipates during rest. In order to test possible group differences in the rate of accumulation of inhibition and to replicate earlier findings, a second experiment, employing a pure massed to spaced performance schedule, was undertaken.

A later experiment, in which high and low satiaters were administered a completely spaced schedule, is also reported here in order to establish that performance differences between satiation groups can be entirely attributed to inhibition effects. It was anticipated that performance differences between groups would not be found under the spaced schedule.

Experiment II

Method

Subjects and procedure.—In selecting *Ss* for the massed to spaced schedule, the KAE was administered twice, in a single session, to 106 female college sophomores who volunteered for the experiment. Approximately half the *Ss* rested 5 min. between administrations; for the remaining *Ss* readministration of the KAE was about 20 min. later. This difference in test-retest time occurred because *Ss* were also participating in another experiment (Spitz & Lipman, 1960). The test-retest tetrachoric r of the KAE was $+.74$ ($P < .001$).

The *Ss* whose KAE scores were consistantly high or consistantly low on both presentations were selected for inclusion in Groups HS and LS, respectively. Three *Ss* scheduled for the rotary pursuit task did not complete the experiment. Apparatus failure was responsible for the elimination of an additional *S*. Group LS consisted of 13 *Ss* whose mean satiation scores were .37 mm. and .10 mm.; Group HS of 16 *Ss* with mean satiation scores of 5.60 mm. and 4.46 mm. There was no overlap of satiation scores among the groups.

In view of the relatively high KAE reliability in the college population, Groups HS-S (high satiation, spaced schedule) and LS-S (low satiation, spaced schedule) were selected on the basis of a single KAE administration which was obtained on 111 female college sophomore volunteers. The *Ss* with the 18 highest and lowest KAE scores were selected. Two *Ss*, 1 HS-S and 1 LS-S, failed to complete the experiment. Group HS-S consisted of 17 *Ss* whose mean satiation score was 6.21 mm.; the mean satiation score for the 17 *Ss* of Group LS-S was .16 mm.

Rotary pursuit schedule.—Group HS and LS received 21 consecutive 20-sec. trials (7 min. pure massed) followed by a 1-min. rest and 21 additional 20-sec. spaced trials with 40-sec. intertrial rest. Continuous, 20-sec. readings were taken during the pure massed trials by employing two standard 6 v. dc timers and a throw switch which *E* operated manually.

Groups HS-S and LS-S received 42 trials of 20-sec. duration and 40-sec. intertrial rest.

Results

The performance curves are shown in Fig. 2. A comparison of Groups HS and LS on Trials 1–21 (lower

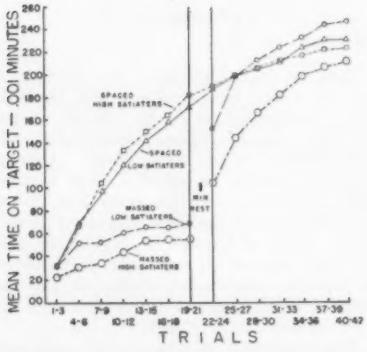

FIG. 2. Mean time on target scores for high and low satiaters under a spaced schedule, and under a massed to spaced schedule. (Trials are plotted in blocks of three. One-minute rest for massed to spaced groups only.)

curves) revealed no reliable ($P > .20$) performance differences. (This finding has since been replicated with an additional 19 low and 22 high satiaters.) The reminiscence gain (the difference between the first three spaced trials and the last three massed trials) was reliably greater in Group LS ($P < .05$) who also performed at a significantly higher level over all spaced trials ($P < .05$) and, of particular interest, over the last three spaced trials ($P < .01$). To control for the somewhat higher first trial performance of Group LS, these data were re-examined based on 10 pairs very closely matched on Trial 1 performance scores. This analysis employed the signed-rank test and confirmed the P values reported above.

An inspection of the upper curves of Fig. 2 reveals that the performance curves for the spaced practice of high and low satiaters are remarkably similar. No reliable performance difference in total time on target over all trials or over the last three trials was found. However, both spaced groups performed significantly ($P < .001$) better over Trials 1–21 than the massed practice groups of Exp. II.

A comparison of Trials 40–42 of the LS-S, HS-S, and LS groups, revealed no significant time on target differences between any of the groups. Although, as previously noted, Group LS performed at a significantly higher asymptotic level than Group HS, it was felt that the most reliable test for the presence of conditioned inhibition in Group HS would be a comparison of their asymptotic performance with that of the combined performance of the three groups (LS-S, HS-S, and LS) in which inhibition decrements could be ruled out. This comparison, over Trials 40–42, resulted in the rejection of the null hypothesis ($P = .04$).

Discussion

The similarity of the performance curves of the high and low satiaters under spaced practice strongly suggests that the performance differences found between high and low satiation groups under massed to spaced practice can safely be attributed to inhibition effects.

The failure to find a reliable performance difference between Groups LS and HS under the pure massed practice schedule, and the parallel nature of this portion of the curves, suggests that the rate of accumulation of inhibition is not an important concomitant of satiation differences. This finding plus the replication of the reminiscence and asymptotic differences between the high and low satiaters suggests, within a one-factor framework, that the rate of dissipation of I_R is the "one-factor" analogue of satiation.

When Trial 1 performance scores of high and low satiaters are pooled across groups (after randomly excluding *Ss* to maintain the same number of retarded and normal *Ss* in both groups) and compared by the U test, the resultant P is significant beyond the .01 level.

Thus, there is a small but statistically reliable starting difference between the high and low satiaters. In a very interesting article, Zeaman and Kaufman (1955) determined that individual differences in starting scores on a motor learning task reflected ". . . generalized conditioned inhibition ($_sI_R$), interpretable as a previously learned tendency to rest during assigned periods of prolonged work" (p. 15). In effect then, the high satiater brings somewhat more of this learned tendency to rest to the experimental situation. Moreover, since the experimental effects were significant with this learned tendency to rest statistically controlled, the finding of an asymptotic difference between Group HS and the combined scores of the other three groups (Fig. 2) indicates that the performance decrement of Group HS at asymptote can be attributed to the building in of additional $_sI_R$. Thus it must be assumed that while the amount of I_R developed by the high and low satiaters under the pure massed schedule (Exp. II) did not differ significantly, the composition of I_R did differ. That is, whereas the I_R developed by low satiaters consisted of only I_R, the I_R developed by high satiaters consisted of both I_R and $_sI_R$. In effect, then, high satiaters (in Exp. I and II) convert I_R to $_sI_R$ whereas low satiaters (in Exp. I andII) build in high levels of I_R without converting I_R to $_sI_R$.

It is theoretically difficult to reconcile the depressed performance of low satiaters relative to the performance of the spaced groups over the massed segments of the performance schedules with the fact that they do not develop $_sI_R$ unless it is assumed with Kimble (1949b) that I_R in addition to being a drive which produces resting also ". . . decreases work out-put, probably by interfering with efficient muscular co-ordination" (p. 501).

An inspection of the spaced segment of the performance curves of the high and low satiaters when switched from massed to spaced practice (Fig. 1 and 2) reveals a level but not a slope difference. That is, after massing, the performance curves of high and low satiaters, over Trials 34–48 in Exp. I and Trials 22–42 in Exp. II, are essentially parallel. The parallel nature of these curves (particularly pronounced over the earlier spaced trials) argues against different rates of dissipation of I_R beyond differences which have occurred during the interpolated rest.

The finding of a significantly larger reminiscence gain in the low satiaters (Exp. I and Exp. II) confirms Meier's (1956) finding of a significant negative correlation of −.47 between kinesthetic satiation and reminiscence following 8 min. of continuous work and a 5-min. rest on the inverted alphabet printing task. (This relatively greater reminiscence gain of low satiaters, as compared with high satiaters, cannot, of course, be generalized to relative dissipation rates of figural aftereffects.) The results of the present study and of Meier's study are inconsistent with Rechtschaffen's (1958) failure to find a statistically reliable association between *visual* satiation and an inverted alphabet printing task. However, recent research (Spitz & Lipman, 1960) indicates that visual and kinesthetic satiation are not significantly related. Therefore, Rechtschaffen's speculation that, ". . . different neural areas were being tapped in the measurement of the two phenomena"(p. 290) would explain his negative finding since the inhibition developed in the inverted alphabet task would seem to be primarily kinesthetic. In this connection, the rotary pursuit task was specifically selected for the present study on the basis of Duncan's (1957) finding that, "Reactive inhibition from massed practice on the rotor seems to be entirely the result of kinesthetic stimulation" (p. 619).

In addition to the modality problem in Rechtschaffen's study, he worked with an "inhibition score" calculated by subtracting the number of letters printed during the last 2½ min. of a 5-min. pure massed condition from a prior tally of the number of letters printed during 2½ min. of an earlier spaced condition (30-sec. work—30-sec. rest). The pres-

ent findings indicate that performance correlates of satiation differences are not revealed in a pure massed or a reasonably spaced schedule.

In order to evaluate possible group differences in the rate of dissipation of I_R during the 1-min. rest in Exp. I and Exp. II it was first necessary, within Hull's framework, to gauge the amount of I_R in the high and low satiaters at the end of massing (Trials 31–33 in Exp. I, Trials 19–21 in Exp. II). In the low satiaters, their performance decrement relative to the spaced group over Trials 31–33 in Exp. I and over Trials 19–21 in Exp. II was labeled I_R. In the high satiaters (Exp. I and II), their asymptotic difference relative to the low satiaters over Trials 46–48 and 40–42, respectively, was subtracted from their performance decrement relative to the spaced group over Trials 31–33 in Exp. I and over Trials 19–21 in Exp. II, and the *remaining* decrement was labeled I_R. The reminiscence gain was then evaluated relative to the inferred amount of I_R in these groups. In both Exp. I and Exp. II the reminiscence gain of the high and low satiaters was judged to represent a very nearly equal proportion of the inferred amount of I_R present at the end of massing.

Within a one-factor framework, support was found for Eysenck's (1955) postulate that the high satiater dissipates I_R more slowly than the low satiater. However, no support was found for a different rate of I_R accumulation or differences in the amount of I_R accumulated.

Finally, then, within Hull's inhibition framework, the most parsimonious statement that can be made regarding satiation as a parameter of pursuit rotor performance is that satiation differences are reflected in the level of I_R at which conversion to ${}_sI_R$ takes place. The crucial difference between the high and low satiater is that the high satiater converts I_R to ${}_sI_R$, whereas the low satiater does not. This correspondence between the Köhler-Wallach satiation construct and Hull's ${}_sI_R$ seems particularly meaningful since both ${}_sI_R$ and satiation are conceptualized as primarily central in locus. Whether ${}_sI_R$ and satiation are identical processes or parallel processes characterized by similar rate constants cannot be answered with the present data.

Summary

In order to test the proposed similarity between cortical satiation (Köhler & Wallach, 1944) and Hull's (1943) inhibition constructs, three groups of high and low satiaters (kinesthetic) were run on the pursuit rotor under varied conditions of work and rest. High satiaters, in Exp. I, performed at a lower level over a partially massed condition (20-sec. work—10-sec. rest), showed a smaller reminiscence gain after a subsequent 1-min. rest, and when switched to a spaced schedule (20-sec. work—60-sec. rest) performed at a lower level and showed an asymptotic decrement over the last three spaced trials. High and low satiaters in Exp. II were given a pure massed schedule (7 min.) followed by a 1-min. rest and then switched to a spaced schedule (20-sec. work—40-sec. rest). Although no significant differences were found under continuous work, the high satiaters again showed a smaller reminiscence gain and an overall and asymptotic decrement relative to the low satiaters on the spaced trials. No differences were found between high and low satiaters on a spaced schedule (20-sec.-work—40-sec. rest). A small, but statistically reliable, starting difference favoring the low satiaters was found when first trial performance scores were pooled across the groups.

These results were interpreted as indicating that the high satiater brings a somewhat stronger learned tendency to rest (${}_sI_R$) to the work task and that, in addition, the high satiater converts I_R to ${}_sI_R$ whereas the low satiater, even after prolonged periods of continuous work, does not convert I_R into ${}_sI_R$.

References

Duncan, C. P. On the similarity between reactive inhibition and neural satiation. *Amer. J. Psychol.*, 1956, **69**, 227–235.

Duncan, C. P. Visual and kinesthetic components of reactive inhibition. *Amer. J. Psychol.*, 1957, **70**, 616–619.

Eysenck, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23. (a)

KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510. (b)

KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain injured. *J. Pers.*, 1952, **21**, 118–148.

KÖHLER, W., & DINNERSTEIN, D. Figural after-effects in kinesthesis. In *Miscellanea psychologica Albert Michotte*. Paris: Librairie Philosophique, 1947. Pp. 196–220.

KÖHLER, W., & FISHBACK, J. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, **40**, 398–410.

KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

MEIER, J. M. Interrelationships among MMPI variables, kinesthetic figural after-effect, and reminiscence in motor learning. Unpublished doctoral dissertation, University of Wisconsin, 1956. (Mic. 18–426, University Microfilms, Ann Arbor, Mich.)

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

SPITZ, H. H., & LIPMAN, R. S. The reliability and intercorrelation of individual differences on visual and kinesthetic figural aftereffects. *Percept. mot. Skills*, 1960, **10**, 159–166.

ZEAMAN, D., & KAUFMAN, H. Individual differences and theory in a motor learning task. *Psychol. Monogr.*, 1955, **69**, (6, Whole No. 391).

(Received August 15, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 476–483

CONFLICT AND THE ORIENTATION REACTION

D. E. BERLYNE[1]

Boston University

Among the many factors that have been shown to influence the strength and direction of exploratory behavior, one particular group of closely related stimulus properties predominates (Berlyne, 1960a). These properties may be called "collative" properties, since they depend on the collation of information from different sources. They include degree of novelty, surprisingness, complexity, and power to induce uncertainty.

Meanwhile, other investigations have shown at least some forms of exploratory behavior to be part of a pervasive complex of psychophysiological processes which mobilize the organism's equipment for gathering information through sensory channels and for taking drastic action. This complex, known as the "orientation reaction" (Berlyne, 1960a, Ch. 4; Sokolov, 1958), involves both transitory jumps and more lasting upward drifts in the level of arousal.

The experiments to be reported were concerned with the third side of the triangle, namely the relations between collative variables and arousal. There is no lack of evidence that the level of arousal and the magnitude of the orientation reaction are affected by collative variables in general. But several logically distinct collative variables have generally been confounded in the relevant studies, just as they are often correlated with one another in naturally occurring situations.

In particular, the present experiments were meant to test the hypothesis that the motivational effects common to collative variables are due to conflict. The bearing of this hypothesis on exploratory and epistemic behavior has been examined from a number of angles by the writer (Berlyne, 1954, 1957b, 1960a, 1960b). A related view has been expressed by Polezhaev (1958, 1959a, 1959b, 1960), who finds that the orientation reaction appears whenever antagonistic activities are simultaneously called forth in the dog.

The hypothesis implies that the intensity of the orientation reaction will increase with degree of conflict. Experiments directed at the effects of conflict on indices of arousal have been scanty. There have been studies of disturbances in behavior due to conflict (e.g., Hovland & Sears, 1938), but these have generally confounded degree of conflict with novelty, surprisingness, and stimulus intensity. In order to test our hypothesis regarding the role of conflict, we must devise a situation in which the degree of conflict can be varied without, as far as possible, being correlated with other variables that might affect arousal. We must also sample arbitrary forms of conflict, since most previous investigators have concentrated on special cases, e.g., conflicts between affective evaluations (e.g., Lanier, 1941) or between vital activities (e.g., Polezhaev, 1958, 1959a, 1959b, 1960), and these may well

[1] The author is greatly indebted to William E. Carr, who assisted at all stages of this series of experiments, and to Paul Weisberg, who assisted in the analysis of the data.

The experiments were carried out at the National Institute of Mental Health.

have special properties apart from those possessed by conflict in general.

In the present experiments, the dependent variable was the galvanic skin response (GSR), recorded by the Fels Dermohmmeter. The GSR is recognized as one of the vegetative components of the orientation reaction, and skin conductance is widely accepted as an index of level of arousal. The GSR results from direct stimulation of points throughout the brain-stem reticular formation in the flaxedilized cat (Bloch & Bonvallet, 1960), and the reticular formation is apparently the part of the nervous system that has most to do with the orientation reaction (Berlyne, 1960a, Ch. 4).

Finger volume was also recorded by means of a plastic capsule enclosing the middle finger of the left hand and connected through a Statham pressure transducer to a Sanborn polygraph recorder. But since most of the fluctuations in the baseline did not appear to correspond to presentations of stimuli, this measure will not be discussed in what follows.

Experiment I

In Exp. I, the situations selected as high-conflict and low-conflict conditions were, respectively, the free-choice and forced-choice situations that had previously proved serviceable in a series of experiments on conflict and reaction time (Berlyne, 1957a). The free-choice situation is one in which two visual stimuli corresponding to different key pressing responses appear at once, and *S* has to respond to either stimulus but not both. In the forced-choice situation, there is simply one stimulus corresponding to one response, so that *S* has no freedom of choice as far as his response is concerned. We speak of "high-conflict" and "low-conflict" conditions, rather than of "conflict" and "nonconflict" conditions, because, even in a forced-choice situation, there will presumably be some degree of conflict between the correct response and other, erroneous response tendencies resulting from stimulus generalization.

Method

If the usual procedure of having a short-lasting stimulus and instructing *S* to respond to its onset had been followed, there would generally have been one GSR per trial, and this GSR would have reflected the joint effects of stimulus onset, stimulus termination, and the motor process (including proprioceptive stimulation consequent on the performance of the response) (Terekhova, 1958). It was therefore decided to have a stimulus pattern lasting 10 sec. and to instruct *Ss* to respond to its *termination*. There were thus generally two GSRs per trial. The second of these must have reflected the joint effects of stimulus termination and response, whereas the first GSR was able to represent the effect of being exposed to stimulus elements associated with incompatible responses, without being contaminated by any motor process. This procedure was called for, not only as a means of increasing the sensitivity of the measure of primary interest, but also as a control. High-conflict trials might possibly have produced more vigorous responses, resulting in GSRs of higher amplitude. If the effects on skin conductance of stimulus onset and response had been allowed to coincide, an invalid conclusion that high-conflict stimulus patterns produce larger GSRs might thus have been favored.

Apparatus.—Apart from the Dermohmmeter and plethysmographic equipment, the apparatus consisted of a stimulus panel, a response unit, and a programer. The stimulus panel bore eight GE 44 6-v. bulbs covered with yellow jewels, two at each of the four corners of a diamond. The panel was painted gray, except for a white fixation spot in the center of the diamond. The response unit included a key that could be moved vertically or horizontally in four directions, corresponding to the corners of the diamond.

The programer operated by means of a punched tape and was set to illuminate combinations of the eight lights in a predetermined sequence, each combination last-

ing 10 sec. and separated by an 11-sec. interval from the next.

Controls.—For low-conflict trials, two lights at one corner of the diamond were illuminated (hereafter referred to as an *adjacent* pattern), whereas, for high-conflict trials, two lights at different corners were illuminated (hereafter referred to as a *dispersed* pattern).

Degree of conflict was thus isolated from stimulus intensity, since two lights were on during trials of both types. The adjacent patterns comprised one inner light and one outer light. It was thus necessary to control for distance from fixation point by arranging for half of the dispersed patterns to comprise two inner lights and the other half to comprise two outer lights.

High-conflict and low-conflict situations did not differ in novelty or surprisingness, since *Ss* were apprised beforehand of the nature of the two kinds of trials and informed that they would be intermingled. Moreover, two sequences were used with equal numbers of *Ss*, each having adjacent patterns where the other had dispersed patterns and vice versa.

There remains the possibility that different magnitudes of GSR may be produced by adjacent and dispersed stimulus patterns as such, regardless of the degree of conflict between overt responses that they occasioned. They may, for example, be regarded as differing in complexity. To test for this possibility, *Ss* were put through a preliminary sequence of adjacent and dispersed patterns in connection with which no overt response was required or even mentioned.

Subjects.—There were 24 *Ss*, 12 men and 12 women. The men were 10 normal-control volunteer inpatients in their late teens or early twenties, one library assistant, and one research assistant in social psychology. The women were eight psychiatric nurses, one typist, one administrative assistant, one librarian, and one research sociologist.

Procedure.—The two sequences, X1 and X2, each comprised 12 adjacent patterns and 12 dispersed patterns in a random order. The 4 possible adjacent patterns were used three times each, and the 12 possible dispersed patterns of two inner or two outer lights once each. Half of the *Ss* in each sex group had X1 for Phase I and X2 for Phase II, and the remainder had the sequences in the opposite order.

The *S* was seated in a dimly lit room in front of the table, facing the fixation spot. All recording and programing equipment was located in an adjoining room, in which the 2 *Es* stationed themselves during the presentation of stimuli. The capsule for measuring finger volume was attached to the middle finger of *S*'s left hand. The GSR electrodes were attached to the soles of the feet, since the right hand was to be used for key pressing, and the record would have been loaded with movement artifacts if the palmar GSR had been measured. There was an interval of 15 min. for skin hydration. Then *S* underwent the two experimental phases.

In Phase I, *S* was told that lights would go on and off. Sometimes there would be two lights at the same corner, and at other times there would be two lights at different corners. He was to do nothing but sit back with his eyes fixed on the fixation spot, trying not to look at the lights directly. He was also to keep his feet and left hand as still as possible, moving them, if necessary, only while no lights were on. The sequence of stimulus patterns was then presented.

In Phase II, *S* was shown the response unit, and it was explained that the key could be moved in any of four directions, corresponding to the four corners of the diamond. There would be a sequence of lights like the one he had just experienced. He was to keep the key grasped between the thumb and first two fingers of the right hand and to move the key as soon as the lights went off. When there were two lights at the same corner, he was to move the key in the direction indicated by the lights, as soon as they disappeared. When there were two lights at different corners, he was to move the key in either, but not both, of the directions indicated by the lights, as soon as the lights disappeared. Once again, he was to keep his eyes on the fixation spot and keep his feet and left hand still. The second sequence was then presented.

Results

For each *S*, three sets of GSRs were examined, namely, (*a*) stimulus-onset GSRs for Phase I, (*b*) stimulus-onset GSRs for Phase II, and (*c*) stimulus-termination-and-response GSRs for Phase II. A GSR was recognized if there was a downward deflection in the skin-resistance tracing during the action of the stimulus pattern in the case of *a* and *b* or during the interval between two successive patterns in the case of *c*. The point at which the first competent deflection occurred was treated as the initial point, and the point of minimum resistance reached by any swing starting with

TABLE 1

EXPERIMENT I: TRANSFORMED GSR SCORES

Phase	Adjacent Patterns (Low Conflict in Phase II)		Dispersed Patterns (High Conflict in Phase II)	
	Mean	*SD*	Mean	*SD*
I	0.806	0.812	0.768	0.748
II: S onset	0.924	0.711	1.120	0.724
II: S termination and R	1.557	0.748	1.613	0.654

a competent deflection was treated as a peak-point.

The resistance values read from the tracing were subjected to the log-conductance-change transformation recommended by Haggard (1949) for data intended for analysis of variance. The actual score used was $\log_{10}[10^9(1/R_P - 1/R_I) + 1]$ where R_P and R_I are the resistances, in ohms, at the peak-point and the initial point, respectively.

For each *S*, a mean adjacent-pattern (low-conflict in Phase II) score and a mean dispersed-pattern (high-conflict in Phase II) score were computed for Phase I stimulus-onset, Phase II stimulus-onset, and Phase II stimulus-termination-and-response GSRs. The means are exhibited in Table 1 and the results of analyses of variance in Table 2.

The comparison of most interest is that between dispersed-pattern (high-conflict) trials and adjacent-pattern (low-conflict) trials for Phase II stimulus-onset GSRs. It will be seen that the mean for the former is higher ($P < .01$), as would be predicted from the hypothesis that the intensity of the orientation reaction increases with conflict. The corresponding comparison for Phase I, in which there were no overt responses, produces a nonsignificant difference in the opposite direction. The difference in Phase II cannot therefore be attributed to intrinsic properties of the stimulus patterns, and we may regard the hypothesis that the magnitude of the orientation reaction increases with degree of

TABLE 2

ANALYSIS OF VARIANCE FOR EXP. I

Source	*df*	Phase I Stimulus Onset		Phase II Stimulus Onset		Phase II Stimulus Termination and Response	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between *Ss*							
Order of sequences (O)	1	1.631	1.51	1.488	1.69	1.932	2.28
Sex groups (S)	1	5.437	5.05*	2.886	3.29	0.700	0.83
O × S	1	0.211	0.20	1.115	1.27	2.235	2.75
Residual	20	1.077	76.93***	0.878	16.88***	0.846	13.02***
Total between *Ss*	23						
Within *Ss*							
Adjacent/Dispersed patterns (A/D)	1	0.017	1.21	0.461	8.87**	0.037	0.57
O × A/D	1	0.177	8.36**	0.566	10.88**	0.581	8.94**
S × A/D	1	0.049	3.50	0.018	0.35	0.015	0.23
O × S × A/D	1	0.014	1.00	0.015	0.29	0.011	0.17
Residual	20	0.014		0.052		0.065	
Total within *Ss*	24						
Grand total	47						

* $P < .05$.
** $P < .01$.
*** $P < .001$.

conflict as confirmed. It may be noted that the stimulus-termination-and-response GSRs in Phase II are higher after high-conflict than after low-conflict trials, but the difference is not significant.

In all three analyses of variance, the interaction between adjacent vs. dispersed patterns and order of sequences is significant. This simply reflects the fact that half of the *S*s had Sequence X1 and the other half Sequence X2 in each phase. Sequence X1 began with two dispersed patterns and X2 with two adjacent patterns, and the first two trials always gave much bigger GSRs than subsequent trials. The sex difference is significant for Phase I but not for Phase II. The male *S*s had a higher mean (1.123 as compared with 0.450) in Phase I than female *S*s; however, it will be recalled that the two sex groups differed in other respects.

The overall mean for Phase II stimulus-termination-and-response GSRs exceeded that for Phase II stimulus-onset GSRs (Ms = 3.170, 2.044; $t = 4.02$, $df = 23$, $P < .001$). And the mean for Phase II stimulus-onset GSRs exceeded that for Phase I stimulus-onset GSRs (Ms = 2.044, 1.574; $t = 2.39$, $df = 23$, $P < .05$). The latter result confirms what Russian experimenters have repeatedly found (e.g., Sokolov 1958, Ch. 3), viz., that the orientation reaction is more pronounced when stimuli have to be discriminated for the guidance of an overt response or, as Russian writers put it, when the stimuli are "signal stimuli."

Experiment II

Experiment II studied different degrees of a more complex kind of conflict, namely response uncertainty in a word association situation. The magnitude of the GSR was once again used as an index of the strength of the orientation reaction.

Laffal (1955) administered 100 stimulus words to 80 *S*s calling for free-association responses. He tabulated the alternative responses given to each stimulus word together with their relative frequencies and calculated, for each stimulus word, a response-uncertainty (*entropy*, H) value according to the usual information-theory formula: $H = -\Sigma_i p_i \log_2 p_i$.

He argued that, since verbal associations are primarily culturally determined, each *S* can be assumed to have possessed a hierarchy of responses to each word, approximating the set of responses given by the population as a whole, with relative strengths approximating their relative inter-*S* frequencies. The inter-*S* response uncertainty thus provides an estimate of the mean intra-*S* response uncertainty. It was found that stimulus words with higher response uncertainties occasioned more "response faults," e.g., failure to recall the response when the stimulus word was presented a second time, exceptionally high latency, misunderstanding.

A stimulus word with which a number of competing verbal responses are associated is, of course, a conflict-inducing stimulus, and, in such circumstances, the information-theory measure of uncertainty can be regarded as a measure of the degree of conflict (Berlyne, 1957b, 1960a). It should be pointed out that it is here a matter of an observer's uncertainty about the response that *S* will make. This must not be confused with the collative stimulus property we have called "power to induce uncertainty," which refers to the generation of uncertainty inside *S* with respect to what will follow, although conflict is involved in both cases and the same information-theoretic measure is applicable.

Method

Material.—Six high-uncertainty and six low-uncertainty words were selected from Laffal's list. All of the words at the upper end of Laffal's list that had threatening connotations and also the word "suck" with its possible oral-erotic significance were omitted. Otherwise, the words selected (shown in Table 3) were the six with the highest response uncertainty and the six with the lowest response uncertainty cited by Laffal. The high-uncertainty words had a mean H of 4.7264 ($SD = 0.32$) and the low-uncertainty words a mean H of 1.2294 ($SD = 1.17$), the means being significantly different at the .001 level ($t = 15.75$, 10 df). It will be noted that there were "emotional" words in both categories.

The 12 words were arranged in a randomized sequence with the restriction that each successive pair of words had a high-uncer-

tainty and a low-uncertainty member. The sequence began with a low-uncertainty word and ended with a high-uncertainty word.

Subjects.—There were 13 men and 7 women, all of whom participated in Exp. II either immediately after Exp. I or after pilot work for Exp. I. Ten men heard the stimulus words in the original sequence mentioned above. The remaining *S*s heard them in the reverse order.

Procedure.—The *S*s were instructed to close their eyes and listen while 1 *E*, sitting in the same room, read out a list of words. They were to respond in each case with the first word that came to mind. The words were read at the rate of one word every 21 sec.

Results

As Table 3 shows, the mean transformed GSR score for the high-uncertainty words, 2.508, was higher ($P < .01$, see Table 4) than the score for the low-uncertainty words, 2.332. The conflict hypothesis may thus be considered to have passed a further test.

TABLE 3

STIMULUS WORDS IN EXP. II

Word	Position in Sequence		*H* (Bits)	Mean GSR Score
	Original	Reversed		
High-Uncertainty Words				
SPIT	12	1	4.9380	2.776
TANGLE	6	7	4.8768	2.344
SOOTHE	10	3	4.6808	2.539
SNEAK	8	5	4.6448	2.680
WRINKLE	2	11	4.6208	2.495
DUMP	3	10	4.5975	2.217
Means			4.7264	2.508
Low-Uncertainty Words				
DWARF	5	8	1.9952	2.340
STAMMER	9	4	1.5355	2.374
STUMBLE	11	2	1.4979	1.916
SHOVE	7	6	0.9337	2.278
GARAGE	1	12	0.8088	2.776
RIP	4	9	0.6238	2.312
Means			1.2294	2.332

TABLE 4

ANALYSIS OF VARIANCE FOR EXP. II

Source	*df*	*MS*	*F*
Between *S*s			
Stimulus order (O)	1	0.675	0.03
Residual	18	26.517	28.85**
Total between *S*s	19		
Within *S*s			
Response uncertainty (*H*)	1	9.793	10.66**
O × *H*	1	0.204	0.22
Residual	18	0.919	
Total within *S*s	20		
Grand total	39		

** $P < .01$.

EXPERIMENT III

Numerous experiments have established that the level of arousal and the magnitude of the orientation reaction are affected by stimulus novelty unconfounded with surprise. Separate components of the orientation reaction have been shown to weaken and finally extinguish altogether when a monotonous stimulus is repeated (Berlyne, 1960a). Sharpless and Jasper (1956) demonstrated similar habituation when they recorded the electrical activity of the reticular formation directly. Whether or not the first stimulus in such repetitious sequences is surprising or not for a human *S* depends on how far he has been told beforehand what to expect. But the stimuli will certainly no longer surprise after a few presentations, and each succeeding stimulus will differ from previous ones only in being less novel.

Similarly, many investigators have shown that the sudden unexpected introduction of an alteration in experimental conditions will intensify or disinhibit the orientation reaction. But we need to ascertain whether arousal is affected by surprisingness unconfounded with novelty. This was the aim of Experiment III.

The conflict hypothesis interprets surprise as a special form of conflict, on the grounds that an expectation will bring out one kind of response and a stimulus disagreeing with the expectation will bring out another kind of response that is incompatible with the first. The fact that conflict independent of surprise has been found to influence the orientation reaction in Exp. I and II favors this interpretation.

Stimuli that were surprising without being

novel were contrived in the following manner. The apparatus was the same as that used in Exp. I. The expectations to be violated were built up both by prior verbal description and by repetition of a temporal pattern. Let us call the two lights at the upper corner of the diamond "A" and the two lights at the right-hand corner "B." For one sequence, U1, *S* was told that A and B would light up in turn. The sequence actually consisted of ABABABABABABABAAABAB. At Position 16 the **A** appeared in place of a B and is thus designated the *surprising* stimulus. The *A* in Position 19, which agreed with the expectation of alternation, was designated the *comparison* stimulus. As for the other sequence, U2, *S* was told beforehand that it would consist of B and A in alternation, but it actually consisted of BABABABABABABA-BABAAA. Here, the **A** in Position 19 was the surprising stimulus and the *A* in Position 16 was the comparison stimulus. The stimuli represented by **A**'s were surprising but not novel, since the sequence had already contained several A's and the last stimulus encountered was, in fact, an A.

All *Ss* had both sequences, one with an overt key pressing response required and one without the overt response, order being counterbalanced. GSR data were recorded and analyzed as in the other two experiments.

Subjects.—There were 5 men and 7 women *Ss*, all of whom participated in Exp. III immediately after Exp. I or after pilot work for Exp. I. No. *S* participated in both Exp. II and III. All *Ss* went through two phases, 3 *Ss* having each of the following orders of treatments (R = "overt response"): U1R-U2NoR, U2R-U1NoR, U1NoR-U2R, U2NoR-U1R.

Procedure.—Stimuli were programed to be on for 3 sec. with 9-sec. intervals between them. Before sequences with overt responses, *Ss* were instructed to respond to the onset of each stimulus pattern by moving the key in the corresponding direction. Before sequences without overt responses, they were told to sit back and do nothing. In both cases, they were to keep their eyes on the fixation spot and to keep their feet still.

After the first phase, *Ss* were told that one of the stimuli had appeared out of place because of a hitch in the working of the equipment but it was hoped that such a hitch would not occur during the next phase.

Results

Mean transformed GSR scores are displayed in Table 5. Analysis of variance (see Table 6) shows the mean for surprising stimuli to be significantly higher ($P < .001$) than the mean for comparison stimuli. Particularly interesting, however, is the fact that the interaction between the surprising vs. comparison stimuli and presence vs. absence of overt response did not even approach signficance. What an interaction means depends, of course, on what kind of transformation is applied to the data. With our logarithmic transformation, the lack of a significant interaction means that we have no reason to believe that the *proportion* by which surprise enhances the rise in conductance is affected by the presence or absence of an overt response.

TABLE 5

MEAN TRANSFORMED GSR SCORES: EXP. III

Condition	Surprising Stimuli		Comparison Stimuli	
	Mean	*SD*	Mean	*SD*
With overt R	1.918	1.056	1.132	1.026
Without overt R	0.739	0.870	0.236	0.624

TABLE 6

ANALYSIS OF VARIANCE FOR EXP. III

Source	*df*	*MS*	*F*
Within *Ss*: Between Phases			
Presence/Absence of overt response (R)	1	12.909	14.72**
Sequences (S)	1	1.832	2.09
Phases (P)	1	0.186	0.21
Residual	9	0.877	
Total	12		
Within *Ss*: Within Phases			
Surprising/Comparison stimuli (St)	1	4.987	24.33***
R × St	1	0.240	1.17
S × St	1	0.025	0.12
P × St	1	0.064	0.31
St × *Ss*	11	0.704	3.43*
Residual	9	0.205	
Total	24		

* $P < .05$.
** $P < .01$.
*** $P < .001$.

Some disturbance is only to be expected when *S* is set to receive a stimulus calling for one sort of overt action and receives instead a stimulus calling for another sort of action. But it is instructive, especially in view of its bearing on intrinsic or "disinterested" exploratory activity, to find that a surprising event is no less arousing when it does not have any implications for *S*'s overt behavior. Several *S*s actually reported that they had not been paying attention during the phase without overt responses and had not noticed that one of the stimuli appeared out of turn.

Once again, we see that GSRs were significantly greater in amplitude when overt responses were performed than when they were not. This confirms further that "signal stimuli" produce greater orientation reactions than stimuli without signal value.

Summary

Three experiments were carried out with the dual aim of studying the influence of collative stimulus properties on the amplitude of the orientation reaction and testing the hypothesis that conflict underlies the motivational effects of collative variables. The GSR was used as an index of the orientation reaction.

In Exp. I, using forced-choice and free-choice reactions, GSR amplitude increased with degree of conflict, unconfounded with novelty, surprisingness, intensity, and distance from fixation point, of stimuli. A preliminary phase with no overt responses showed that the effect was not due to inherent properties of the stimulus patterns, e.g., differences in complexity. In Exp. II, using word association, GSR amplitude increased with response uncertainty, regarded as a measure of degree of conflict. In Exp. III, stimuli that were surprising without being novel, surprise being interpreted as a form of conflict, likewise produced more intense GSRs than others.

References

Berlyne, D. E. A theory of human curiosity. *Brit. J. Psychol.*, 1954, **45**, 180–191.

Berlyne, D. E. Conflict and choice time. *Brit. J. Psychol.*, 1957, **48**, 106–118. (a)

Berlyne, D. E. Uncertainty and conflict: A point of contact between information-theory and behavior-theory concepts. *Psychol. Rev.*, 1957, **64**, 329–339. (b)

Berlyne, D. E. *Conflict, arousal and curiosity.* New York: McGraw-Hill, 1960. (a)

Berlyne, D. E. Les équivalences psychologiques et les notions quantitatives. In D. E. Berlyne & J. Piaget, *Théorie du comportement et opérations.* (*Etudes d'épistémologie génétique*, No. 12) Paris: Presses Univer. France, 1960. (b)

Bloch, V., & Bonvallet, M. Le déclenchement des réponses électrodermales à partir du système réticulaire facilitateur. *J. Physiol., Paris*, 1960, **52**, 25–26.

Haggard, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39**, 378–392.

Hovland, C. I., & Sears, R. R. Experiments on motor conflict: I. Types of conflict and their modes of resolution. *J. exp. Psychol.*, 1938, **23**, 477–493.

Laffal, J. Response faults in word association as a function of response entropy. *J. abnorm. soc. Psychol.*, 1955, **50**, 265–270.

Lanier, L. H. An experimental study of "affective conflict." *J. Psychol.*, 1941, **11**, 199–217.

Polezhaev, E. F. [The conditioned postural reflex as a special reaction during the interval between presentations of stimuli.] *Dok. Akad. Nauk SSSR*, 1958, **123**, 204–207.

Polezhaev, E. F. [On cortical co-ordination during the transition between sleep and wakefulness.] *Dok. Akad. Nauk SSSR*, 1959, **126**, 1149–1152. (a)

Polezhaev, E. F. [Novelty as a stimulus for special reactions.] *Byull. eksp. Biol. Med.*, 1959, **2**, 9–14. (b)

Polezhaev, E. F. [Phase changes in EEG as indicators of the formation of cortical co-ordinations.] *Fiziol. Zh. SSSR*, 1960, **46**, 26–36.

Sharpless, S., & Jasper, H. H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655–680.

Sokolov, E. N. *Vospriiate i uslovny refleks.* [Perception and the conditioned reflex.] Moscow: Moscow Univer. Press, 1958.

Terekhova, O. P. [The orientation reflex in conditioned reactions to complex stimuli.] *Vop. Psikhol.*, 1958, **4**(1), 87–96.

(Received August 19, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 484–495

SOME RELATIONS BETWEEN THE INTENSIVE PROPERTIES OF THE CONSUMMATORY RESPONSE AND REINFORCEMENT [1]

GEORGE COLLIER, FREDERICK A. KNARR, AND MELVIN H. MARX

University of Missouri

The present experiment is part of a series (Collier & Myers, 1961; Knarr & Collier, 1960) by which it is hoped to partial out the contributions of the various intensive properties of the consummatory response to reinforcement magnitude. By using a drinking tube in the runway as a magazine, various aspects of the evoking stimuli and the consummatory response can be manipulated. For example, by using solutions varying in quality or concentration the evoking stimuli can be varied. By varying the tube size and by removing the animal from the goalbox after various durations of drinking, numbers of licks, or volumes drunk, these factors can be varied and their effect evaluated.

In the present study the duration of the contact with the magazine was held constant and made sufficiently short so that licking was continuous. Sucrose solutions of 4% and 32% were used. The concentration variable is known to affect the level of instrumental performance, and it is presumed to affect the associated consummatory performance. The present study provides an opportunity to examine their conjoint variation in a situation where postingestive and nutritive factors play a minimal role because of the brief magazine contact and small volumes consumed. Two additional factors which have been hypothesized as determinants of the magnitude of consummatory responding, delay in reinforcement and shifts in reinforcement, have been varied. The manipulation of these variables provides the opportunity of observing the conjoint changes in consummatory and instrumental responses as a function of changes in the character of reinforcement.

[1] This investigation was supported by Research Grants M-817 and M-3328 from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland. A report of these data was made at the 1959 Midwestern Psychological Association meetings in Chicago.

METHOD

Subjects.—The *S*s were 26 female, 150-day-old albino rats from the University of Missouri colony. Two *S*s were discarded for failure on three consecutive trials to traverse the runway in 5 min., and a third because of a foot injury.

Apparatus.—The runway (Fig. 1) was 3½ in. wide by 5 in. high and 18 ft. long; the last 2 ft. were not used. It was housed in a light-tight, sound deadened hull. The *S*s were run in the dark. The floor consisted of nine 2-ft. grid sections of 1½ in. aluminum angle separated end to end by $\frac{3}{16}$ in., and side to side by ¼ in. Section 1 was divided into a 1½-ft. startbox and a 6-in. starttime segment. The next five sections constituted the runway proper. Section 7 contained a 2-ft. delay chamber. This was offset 6 in. so that the last 6-in. part of Section 6 was included in the delay chamber and the last 6-in. part of Section 7 was used to measure starttime out of the delay-chamber segment. The next 2½ ft. contained a 24-in. endbox. Eight times were measured, by means of electronic switches and .01-sec. Standard Electric timers: the time to traverse the first 6 in. out of the startbox, the times to traverse each of the next five 2-ft. sections, the time to traverse the first 6 in. out of the delay chamber, and finally, the time to traverse

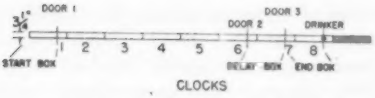

FIG. 1. Schematic of runway.

the endbox and contact the drinking tube. Time spent in retracing was included in the time to traverse the last section reached before the retracing began.

The startbox and the two delaybox doors were operated automatically. The startbox door opened 3 sec. following the closing of the startbox lid. The front delaybox door was down when *S* ran into the delay chamber and opened after a predetermined interval measured from the time *S* touched the seventh grid. The rear delaybox door came down when *S* touched the seventh grid, that is, 6 in. after he entered the delay chamber. Guillotine type doors operated by a motor driven screw were used.

A 50-ml. burette with a 4-mm. opening was used to deliver the reinforcement. It was located behind a 1-in. circular opening 2½ in. above the floor in a partition in the endbox. The first contact with the tube started a Standard Electric timer. All subsequent licks were counted.

Pretraining.—One week before magazine training *Ss* were put on a 22-hr. food and water deprivation schedule and were randomly assigned to the two concentration groups of 4% and 32% sucrose (percentage by weight).

The *Ss* were adapted to *E* and the apparatus for 3 days just prior to the regular feeding time. Each *S* was handled 4 min. per day, then given a 5-min. exploration period in the runway. During exploration all doors were open.

Magazine training.—Each concentration group was randomly divided into two subgroups of 6 *Ss* each. One subgroup of each concentration was trained and tested on alternate days. All *Ss* were trained for 8 days to drink from a burette containing the appropriate solution.

Runway training.—Each *S* was given 2 trials per day in the runway on 21 alternate days. The delay door was down, forcing *S* to slow down or stop, and opened 1 sec. after *S* touched Section 7. The time allowed to drink was held constant at 20 sec. for 12 consecutive trials, 40 sec. for the next 8 trials, and 1 min. for the remaining 11 days of training and the remainder of the experiment. The time from the removal of *S* from the goalbox until he was placed again in the startbox, for his second daily run, was approximately 40 sec.

Delay.—The 4% and 32% *Ss* were each matched into three delay groups of 4 each on the basis of their total running spe during the last six pretest trials. Three delays were used: 1 sec. (same as training), 3 sec., and 45 sec.

Two test trials per day were given each *S* on 3 alternate days. Half of the *Ss* in each of the six groups were tested on different days; however, all *Ss* continued running on the same alternate days as in training.

Retraining.—One week following the delay test trials *Ss* were retrained to asymptotic runway speeds on 8 alternate days. Retraining conditions were identical to the pretest runway conditions except that the 1-sec. delay was eliminated.

Shift.—Two test trials were given each *S* on 10 alternate days. The *Ss* for each concentration were matched into two groups on the basis of their total running speed during the last six retraining trials. The first group of each concentration was shifted to an 11.3% sucrose solution, whereas the second group was shifted to the opposite concentration (i.e., 4 to 32% and 32 to 4%).

RESULTS

Training

Total speeds of running (over the 10½ ft. from startbox door to delaybox) for the means of the 21 pairs of daily trials for the 4% and 32% concentration groups are presented in Fig. 2. An analysis of variance of these data (Table 1) shows that both the terminal speed and the rate of approach (Concentration × Trials interaction) to these speeds are func-

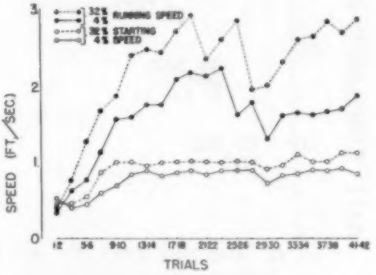

FIG. 2. Running speed (Clocks 1–6) and starting speed out of the delaybox (Clock 7) averaged over daily pairs of trials for the first 42 trials.

TABLE 1

ANALYSES OF VARIANCE OF TOTAL RUNNING SPEED (CLOCKS 1–6) AND TUBE CONTACT SPEED (CLOCKS 7–8)

Source	*df*	Running Speed		Tube Contact Speed	
		MS	*F*	*MS*	*F*
Between *Ss*	23				
Concentration (C)	1	100.47	6.60*	7.63	7.43*
Error (b)	22	15.23		1.03	
Within *Ss*	984	.70		.07	
Trials (T)	41	7.86	21.81**	.79	21.05**
C × T	41	.92	2.55**	.06	1.68*
Error (w)	902	.36		.04	

* $P < .05$.
** $P < .01$.

tions of concentration. It should be noted that this level of running is closely approached by Trial 12 of the 32% group, i.e., over those trials in which only 20 sec. contact time on the drinker was allowed.

In Fig. 2 is also presented the speed determined from the opening of the delaybox door (1-sec. delay) to the first contact with the drinking tube. An analysis (Table 1) of these data yielded results similar to those for total speeds.

In the first portions of Fig. 3 are presented the number of licks and the amount consumed for the two concentration groups. Analyses of these

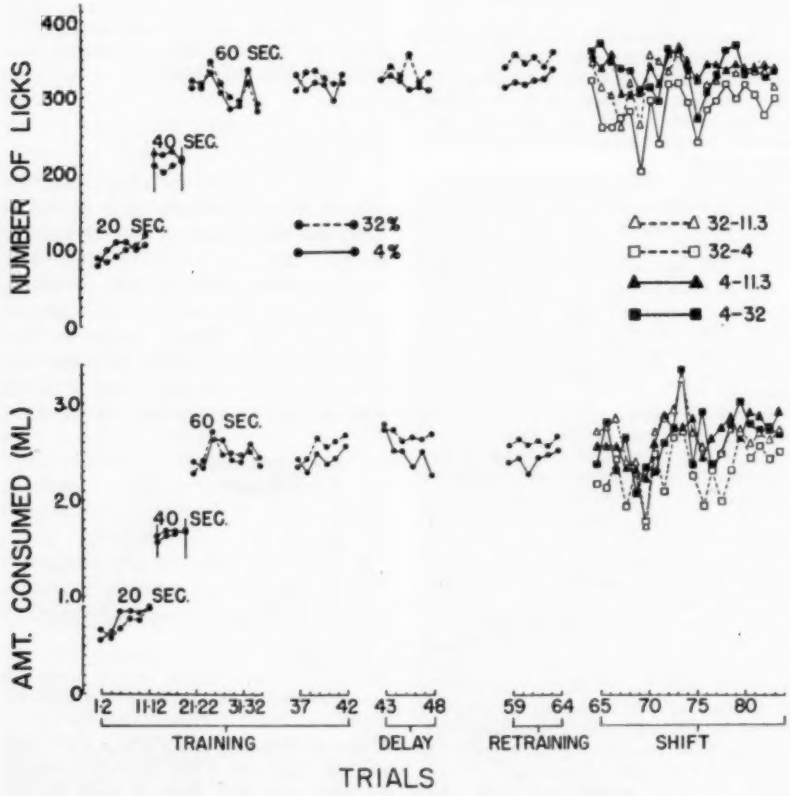

FIG. 3. Amount consumed and number of licks over the course of the experiment.

consummatory measures for the last eight trials at each drinking time (i.e., 20, 40, 60 sec. blocks) revealed no significant differences for trials within blocks, concentration, or their combination. Although in the early trials the 4% *Ss* characteristically showed a greater mean number of licks. The differences between the numbers of licks and between the volumes consumed at the different drinking times (20, 40, and 60 sec.) were significant. Thus large differences in the rate of acquisition and terminal level of starting and running speed occurred with no discernible differences in total amount of consummatory activity. It can be seen from Fig. 3 that *Ss* consumed approximately 2.5 ml. in 325 licks in 60 sec. of contact with the magazine. Any discussion of relative lick rates or ingestion rates of sugars of different concentrations must take the differences of viscosity into account. At 20° C., 32% sucrose is a little better than three times as viscous as 4% sucrose (Swindells, Snyder, Hardy, & Golden, 1958). This variable might account for the failure of the 32% groups to show a greater number of licks in the early trials. The data presented later in the present report show a progressive increase in the number of licks and volume consumed in 60 sec. so that the 32% groups eventually significantly exceeded the 4% on these two measures. Because of the differences in viscosity, lick rate and ingestion rate are not equivalent over a range of concentrations.

Delay

In order to assess the effect of delay the following analyses were performed on the six trials preceding the introduction of delay and the six delay trials.

Table 2 presents the mean starting speeds out of the startbox and the delay chamber (Clocks 1 and 7) for the three delay groups. It is clear from the analysis of variance of these means in Table 3 that the major effect of the introduction of the delay was on the delaybox starting times and produced a reduction in starting speed which was a function of the length of delay. The failure of delay to affect startbox starting speeds is to be expected since it usually has

TABLE 2

STARTING SPEEDS (FT/SEC) OUT OF THE STARTBOX (CLOCK 1) AND THE DELAYBOX (CLOCK 7) AS A FUNCTION OF DELAY

	Startbox			Delaybox		
	1 sec.	3 sec.	45 sec.	1 sec.	3 sec.	45 sec.
Predelay	.36	.34	.36	.41	.45	.42
Delay	.35	.35	.35	.39	.41	.24

TABLE 3

ANALYSIS OF VARIANCE OF SPEEDS FOR CLOCKS 1 AND 7

Source	*df*	*MS*	*F*
Between *Ss*	23		
Concentration (C)	1	117.46	12.07**
Delays (D)	2	9.85	.32
C × D	2	20.90	.67
Error (b)	18	31.16	
Within *Ss*	552	853.79	
Pretest-Test (E)	1	20.66	11.50**
Trials (T)	5	.60	.53
Clocks (V)	1	14.20	2.67
E × T	5	.42	.40
E × V	1	5.84	9.09**
E × C	1	23.35	2.02
E × D	2	10.42	5.80*
.			
.			
.			
D × E × V	2	7.69	11.99**
.			
.			
.			
C × D × E × T × V	10	1.26	1.02
Error (w)	414	1.28	
e_1 (CDES)	18	1.80	
e_2 (CDTS)	90	1.14	
e_3 (CDVS)	18	5.31	
e_4 (CDETS)	90	1.02	
e_5 (CDEVS)	18	.64	
e_6 (CDTVS)	90	.95	
e_7 (CDETVS)	90	1.23	

* $P < .05$.
** $P < .01$.

TABLE 4

ANALYSIS OF VARIANCE OF SPEEDS FOR CLOCKS 2–6

Source	*df*	*MS*	*F*
Between *Ss*	23		
Concentration (C)	1	786.50	12.07**
Delays (D)	2	13.74	.21
C × D	2	6.39	.10
Error (b)	18	65.17	
Within *Ss*	1416	2.52	
Pretest-Test (E)	1	15.72	2.50
Trials (T)	5	7.51	2.56*
Clocks (V)	4	71.25	31.24**
.			
.			
.			
V × C	4	17.23	7.56**
.			
.			
.			
C × E × T	5	7.87	3.01*
.			
.			
.			
C × D × E × T × V	40	1.95	1.01
Error (w)	1062	2.24	
e_1 (CDES)	18	6.28	
e_2 (CDTS)	90	2.93	
e_3 (CDVS)	72	2.28	
e_4 (CDETS)	90	2.61	
e_5 (CDEVS)	72	2.23	
e_6 (CDTVS)	360	2.08	
e_7 (CDETVS)	360	1.94	

* $P < .05$.
** $P < .01$.

its effect over trials. Concentration did not interact with any of the other variables.

Considering now the running speed between the startbox and the delaybox (Clocks 2, 3, 4, 5, and 6), neither the main effect of delay nor any of its interactions was significant (Table 4). Figure 4 shows the gradients of running, averaged over the two (pre- and postdelay) blocks of six trials.

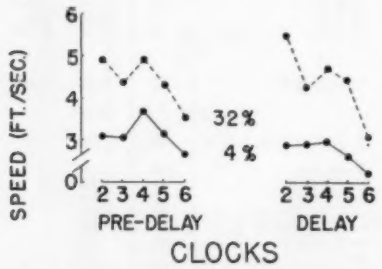

FIG. 4. Gradient of running to the delaybox before and after the introduction of delay.

TABLE 5

SPEED OF APPROACH (FT/SEC) TO DRINKER AS A FUNCTION OF DELAY AND CONCENTRATION

Cond.	Trial	Delay					
		1 sec.		3 sec.		45 sec.	
		4%	32%	4%	32%	4%	32%
Predelay	6	1.14	1.95	1.50	2.98	1.62	2.95
Delay	1	1.46	2.33	1.89	2.45	1.00	1.09
	6	1.40	2.26	2.01	1.79	1.05	1.69

The fastest running occurred early in the runway. The gradient was steepest for the highest concentration. Again as with starting speed the failure of delay to influence the speed of approach to the delaybox is to be expected because of the small number of delay trials. It is interesting to note in Fig. 4 that the introduction of delay did not produce any immediate effect on the shape of the gradient of approach.

The major effect of the introduction of delay on the speed of approaching the drinker (Clock 8), as shown in Table 5, was to produce an *immediate* (first delay trial) drop in the speed of leaving the delaybox. An analysis of variance of these data is shown in Table 6. These results parallel those on the starting speeds from the delaybox (see Table 2). The size of this drop was an increasing function of the length of delay. Further, the greatest drop occurred for the 32% group, which showed the significantly greater speed, with little or no drop occurring for the 4% group. These data parallel those found by Collier and Marx (1957) for a delay introduced in the center of this runway.

Analyses of variance, (Table 7) of the volume consumed and of number of licks showed the interactions of pre- and postdelay blocks and trials and of trials and concentrations sig-

TABLE 6

ANALYSIS OF VARIANCE OF SPEEDS FOR CLOCK 8

Source	*df*	*MS*	*F*
Between *Ss*	23		
Concentration (C)	1	25.97	11.90**
Delays (D)	2	7.43	3.40
C × D	2	.05	.02
Error (b)	18	2.18	
Within *Ss*	264	.35	
Pretest-Test (E)	1	2.56	4.33
Trials (T)	5	.66	2.95*
E × D	2	7.44	12.57***
C × E × T	5	.59	2.67*
D × E × T	10	.28	1.28
C × D × E × T	10	.42	1.91
Error (w)	198	.26	
e_1 (CDES)	18	.59	
e_2 (CDTS)	90	.22	
e_3 (CDETS)	90	.22	

* $P < .05$.
** $P < .01$.
*** $P < .001$.

nificant for volume consumed (see Fig. 3). Both of these interactions appear to have resulted from a tendency, increasing over trials, for the 32% *Ss* to consume more solution within the 1-min. drinking period. Delay appears to have affected the 4% groups, in that they showed a tendency to decline in consummatory activity following the introduction of delay (see Fig. 3). This pattern was reproduced in the lick data (Fig. 3), but did not reach significance.

TABLE 7

ANALYSIS OF VARIANCE OF VOLUME CONSUMED AND NUMBER OF LICKS

Source	*df*	Volume Consumed		No. of Licks	
		MS	*F*	*MS*	*F*
Between *Ss*	23				
Concentration (C)	1	1.68	1.00	10,841.3	.79
Delays (D)	2	.43	.26	1,838.9	.13
C × D	2	1.17	.70	22,875.1	1.66
Error (b)	18	1.66		13,755.2	
Within *Ss*	264	.09		1,306.6	
Pretest-Test (E)	1	.82	6.47*	4,087.6	1.38
Trials (T)	5	.07	1.08	2,057.5	1.89
E × T	5	.40	4.53**	1,138.5	.92
T × C	5	.17	2.67*	892.4	.82
C × D × E × T	10	.05	.51	1,019.6	.83
Error (w)	198	.08		1,325.1	
e_1 (CDES)	18	.13		2,960.6	
e_2 (CDTS)	90	.06		1,089.7	
e_3 (CDETS)	90	.09		1,233.2	

* $P < .05$.
** $P < .01$.

Retraining

The data of the last 3 of the 8 retraining days (i.e., six trials) were analyzed. The starting speeds (Clock 1) differed only with respect to concentration. These times were, for 4% and 32%, respectively, .26 and .47 ft/sec. The analysis of running speed for Clocks 2–7 (with delay chamber now removed) showed Clocks ($P \leq .01$; $F_V = 32.26$; $df = 5/110$), Concentration ($P \leq .01$; $F_C = 71.84$; $df = 1/22$), and their interaction ($P \leq .01$; $F_{VC} = 27.51$; $df = 5/110$) to be significant. The differences between Trial 1 and Trial 2 on any given day, between days, and their interactions proved to be insignificant. A plot of the Concentration

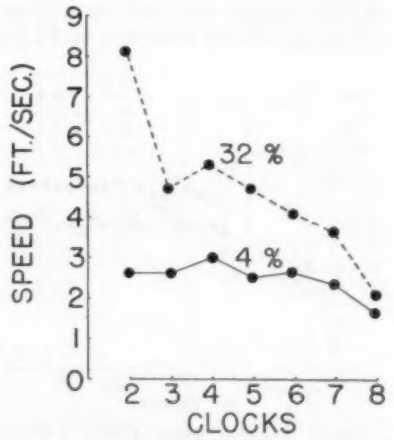

FIG. 5. The gradient of running to the endbox averaged over the last six retraining trials.

× Clocks interaction (including Clock 8) in Fig. 5 shows a steep gradient of running for the 32% *Ss*, the fastest running occurring in the initial portion of the runway; and little or no gradient for the 4% group. This curve is similar to those shown in Fig. 4. The additional trials appear to have increased overall speed and steepened the gradient. An analysis of the drinker-approach speeds (Clock 8) showed concentration to be significant ($P \leq .01$; $F_C = 8.52$; $df = 1/22$). The analysis of volume consumed showed no differences (Fig. 3), while the analysis of number of licks showed a significantly larger number on the second trial of each day ($P \leq .01$; $F = 10.98$; $df = 1/10$), and a significantly greater number of licks for the 32% *Ss* (Fig. 3) ($P \leq .01$; $F = 11.97$; $df = 1/22$).

Shift

Figure 6 details the effect over trials of the shifts in concentration of reinforcement on *total* running speed. Certain relations are readily apparent. By the third trial following shift, *Ss* changed from 4% to 32% sucrose reached approximately three-fourths of their average terminal rate whereas the group shifted from 4% to 11.3% reached about four-fifths of their final rate. On the other hand, *Ss* shifted from 32% to 4% or 11.3% showed a gradual decline in running speed and an interesting periodicity. The second trial of each day was performed at a lower speed than the first trial and the first trial of each day was higher than the last trial of the preceding day, with very few exceptions. This periodicity was most pronounced for *Ss* undergoing the maximum downward shift (32% to 4%). The terminal rate for both groups of *Ss* shifted to 11.3% was the same, approximately halfway between the preshift rates for the speeds on 4% and 32%. The *Ss* shifted from 4% to 32% reached an asymptotic terminal speed above the 11.3% animals and below the preshift 32% speed. Periodicity, of the sort found for the shift-down groups, did not occur significantly in the total speed measures. The *Ss* shifted from 32% to 4% arrived at a level below the 11.3% groups but above the preshift 4% animals. It is not clear whether this last group was asymptotic.

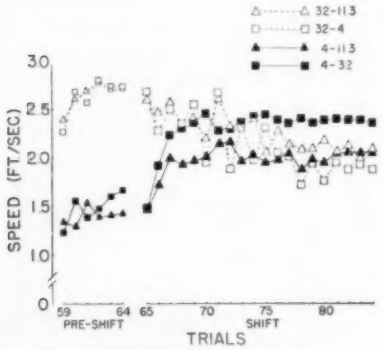

FIG. 6. Running speed (Clocks 1–8) on the six trials preceding and the 20 trials following shift in reinforcement.

In order to gain a more detailed look at these data, including the gradient of running, a series of analyses of various combinations of the shift groups were performed. The immediately preceding six *preshift* trials, the first six, and last six *shift* trials, the segments of the runway, and the first and second trials of each day were considered.

Comparisons of the 4% and 32% groups shifted to 11.3%.—Figure 7 presents the data for the groups shifted to 11.3%. Table 8 presents an analysis of these data for Clocks 2–7.[2] The first thing to be noted is

[2] The same variables as in the preceding analysis with the exception of the concentration groups were examined in the subsequent analyses. In order to save space

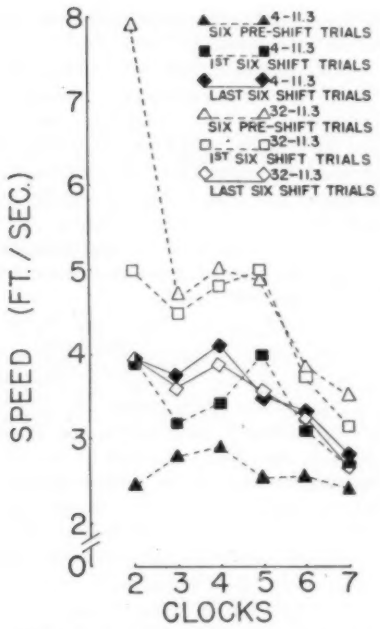

FIG. 7. The gradients of running on the six trials preceding, the six trials immediately following, and the final six trials following the shift in reinforcement.

that the steepness of the running gradient over the length of the runway was an increasing function, with the 11.3% gradient falling between the flat 4% gradient and the steep 32% gradient. As was the case with the shift in total speed, the 4%–11.3% group showed an almost immediate shift in gradient, while the 32%–11.3% group showed a slow shift to the 11.3% gradient. Most of the shift occurred in the last block of six trials. When Clock 1 and Clocks 2–7 were examined separately for the two shift-down groups (32%–11.3% and 32%–4%), it was found that the major part of the trial-to-trial periodicity shown in the total running speed measures was the result of starting speed measures rather than running speed measures. Rats started faster on the first trial than the second, but once running, ran at approximately the same speed on both of the two daily trials. The analysis on the speed over the last runway segment (Clock 8) indicates that the shift-up group's approach to the drinker was little affected by the shift, while the shift-down group showed a gradual decline to a slower approach to the drinker.

TABLE 8

ANALYSIS OF VARIANCE OF RUNNING SPEED OF THE 4%–11.3% AND 32%–11.3% SHIFT GROUPS

Source	*df*	*MS*	*F*
Between *Ss*	11	72.76	
Concentration (C)	1	400.37	10.00*
Error (b)	10	40.00	
Within *Ss*	1284	2.64	
Days (D)	2	.53	.34
Trials/Day (T)	1	2.78	1.27
Clocks (V)	5	76.07	20.64**
Pretest, $Posttest_1$, $Posttest_2$ (E)	2	15.36	2.61
.			
.			
.			
C × E	2	162.81	27.64**
C × V	5	20.92	5.66**
.			
.			
.			
C × V × E	10	13.18	3.86**
.			
.			
.			
C × D × V × E	20	1.93	2.05**
.			
.			
.			
C × D × T × V × E	20	1.22	.59
Error (w)	1070	1.93	
e_1 (CDS)	20	1.56	
e_2 (CTS)	10	2.18	
e_3 (CVS)	50	3.69	
e_4 (CES)	20	5.89	
.			
.			
.			
e_{10} (CVES)	100	3.41	
.			
.			
.			
e_{14} (CDVES)	200	.94	
e_{15} (CDTVES)	200	2.06	

* $P < .05$.
** $P < .01$.

neither the tables nor the significant effects are mentioned. No nonsignificant effect is discussed without this fact being noted. All significant effects are mentioned. Summaries of the designs may be had on request.

Comparisons of the preshift and postshift 4% and 32% groups.—Two analyses, one comparing the preshift 4% group with the postshift 4% (i.e., 32% to 4%) groups, and the second comparing the preshift 32% group with the postshift 32% groups (4%–32%), yielded some interesting comparisons. In both cases, the starting speed (Clock 1) was significantly different. The *S*s shifted from 32% to 4% never reached the slower starting speeds of the original 4% *S*s. Similarly, the *S*s shifted from 4% to 32% never obtained the speed of the original 32% *S*s. The shifted *S*s again showed the significant periodicity only on Clock 1. Considering the gradient of running (Clocks 2–7), neither of the 4% groups, the original or the postshift, showed a significant gradient, while both of the 32% groups did. Here again the *S*s shifted to 32% never obtained the level of speed of the original group nor did they show the steepness of gradient of running.

Comparison of the 32% groups shifted to 4% and 11.3% and of the 4% groups shifted to 11.3% and 32%.—An analysis of the two groups shifted from 32% (to 11.3% and 4%) showed that they differed significantly in the final gradients (Clocks 2–7) and in the rate of approach to their final gradient but that they did not differ significantly on Clock 8, although they both were significantly slower than their original speed. Periodicity was again significant only for starting speed. An analysis of the two groups shifted from 4% (to 11.3% and 32%) showed that they differed significantly in their final level, that on starting speeds those *S*s shifted to 32% ran faster on the second trial of each day's run while those shifted to 11.3% did not show this periodicity, and that the gradient of running was greater for both groups following the shift.

Consummatory behavior.—An analysis of the consummatory measures, licks, and amount consumed, on the last six trials of retraining (see Fig. 3) showed no significant differences for amount consumed, but a significant difference in number of licks. The 32% group showed a significantly greater number of licks on the second of the two daily trials.

An analysis of the two 11.3% groups, those *S*s shifted from 4% and those *S*s shifted from 32%, showed no significant differences in either licks or amount consumed. An analysis of the 32% *S*s shifted to 4% and those shifted to 11.3% showed no terminal differences in amount consumed or number of licks. Both measures showed a drop in the first block of six shift trials with recovery by the last block of six trials. Both measures also showed a significant tendency for less solution to be drunk with fewer licks in the first shift trials. An analysis of the 4% *S*s shifted to 11.3% and 32% showed that both groups consumed and licked more following the shift with greatest amount occurring by the last block of trials. Both groups showed a significant tendency to drink more and lick more on the second of the daily trials, with *S*s shifted to 32% drinking significantly more than those shifted to 11.3%.

Analyses comparing the preshift and shifted 32% groups showed only a significant increase in the number of licks on the second of the two daily trials. A similar comparison of the preshift and shifted 4% groups showed a significant tendency for the 4% *S*s shifted from 32% to make fewer licks and get more solution than the original 4% group. Further, the postshift 4% group consumed sig-

nificantly more on the second of the daily trials than did their predecessors.

Discussion

If the consummatory response and its ancillaries are to be considered "responsible" for variations in performance of an instrumental response, as some accounts of reinforcement suggest (e.g., Estes, 1954; Premack, 1959; Spence, 1956), there must be some characteristics of the consummatory response, qualitative or quantitative, which are functions of the variables known to manipulate performance. In the present experiment two properties of the consummatory response, total frequency of licks and total amount consumed per trial, were examined as a function of the intensity (concentration) of the reinforcing stimulus. It was found that after a large number of trials both measures differed significantly as a function of the intensity of the reinforcing stimulus. Over a 1-min. interval of drinking *Ss* licked faster and drank more of the 32% sucrose when compared with 4%. However, it was found that differences in starting speed and running speed evinced themselves very early, long before these differences in lick rate or amount consumed appeared,[3] and in fact the early differences that did appear were in the "wrong" direction. If then the technique of measurement of these variables can be considered to produce reliable results, it can be concluded that neither variations in the amount consumed nor the number of licks in 1 min. of reinforcement are *necessary* conditions for the production of differences in the instrumental response. This conclusion is reinforced by studies in which these variables are directly manipulated by varying tube size and amount consumed jointly (Knarr & Collier, 1960) or number of licks (Goodrich, 1960). It may be argued though that it is only differences in initial consummatory rate that are important. The results of the early trials in training require that this initial rate be sampled for less than 20 sec. It may not be concluded from these results that there is no property of the consummatory response which is necessary for differences in performance, but they seriously limit the possibilities.

The model of runway behavior which views it as a series of discrete responses pinned together by stimuli arising from the runway, from the deprivation operation, and from fragments of the consummatory response occurring to the preceding stimuli, leads naturally to the expectation of a gradient of running. A response chain is constructed from the terminal end, the responses most remote from the reinforcing stimuli being least strengthened. The present results, obtained under conditions where extraneous stimuli were minimized, show a gradient contrary to the "goal gradient" expectations. While no data were obtained for the very early trials, the picture obtained is of the evolution of a gradient with the fastest running occurring earliest in the runway. The steepness of this gradient is a function of variables which manipulate the effectiveness of the reinforcement, for example, shift in reinforcement magnitude. The failure to obtain a gradient accelerated in the direction of the endbox, under conditions where external stimuli are minimal, raises questions about the serial model previously employed to account for runway behavior. Functionally there appear to be only three units—starting, running, and stopping—which can be independently manipulated. This is shown by the effects of delays, shifts, etc., upon these measures. Starting appears to be a function of local stimuli, both extraneous and constant, and of the consequences of the chain. Starting appears to be, from these and prior data (Collier & Marx, 1957; Wist & Collier, 1960), more sensitive than running to amount, changes, and delay of reinforcement. An alternative to the

[3] The early differences were in the direction of slower drinking for the higher concentrations, perhaps reflecting viscosity differences. The development of a superior rate of drinking of 32% sucrose was relatively slow, considering the 8 days of preliminary magazine training.

classical view of running would be that it is a repetitive cyclic activity which is not chained by external stimuli but which is controlled by the terminal discriminative and reinforcing stimuli. It can be broken into segments by bringing it, at various points, under discriminative control, as is shown in maze studies or by the introduction of delays or other obstacles. Studies by Collier and Marx (1957) and Wist and Collier (1960) on the gradient of running preceding and following a point of delay have shown that such an obstacle has its effect mainly on the behavior immediately preceding and following the point and on starting behavior. Once running starts, it occurs at its usual (predelay) rate. Stopping appears to be a simple matter of preservation and to be under the discriminative control of the stimuli preceding the block or end of the runway.

The introduction of a delay results in an abrupt decrement in the performance in the running behavior following the delay point and a gradual decline over trials in the speed of running up to the delay point (Collier & Marx, 1957). It is apparent from the present results that there is no accompanying change in performance of the consummatory response as a function of the kind or durations of delay used, and that, therefore, disturbances in the consummatory response or its precursors can not be used to account for the disturbances in running.

The terminal level of responding and the course of the change are the major data in shift studies. No evidence for the "overshooting" of the preshift asymptotes as a function of shift in reinforcement was obtained in the present study. Rather, the residual effects of the rat's previous history were inertial. Both 11.3% groups, one shifted from 32% and one from 4%, reached a final common asymptote, whereas the 4% group shifted up to 32% reached a stable level below the previous 32% *Ss* and the 4% group shifted from 32% reached a level of responding at the end of 20 trials above that for the original 4% group. These remarks must be qualified by noting that no control groups, i.e., groups continued on 4% and 32%, are available for comparison. It is possible that the speed of running associated with these values of reinforcement converge, in which case the performance might meet or exceed the control performance. Such a substantial drift in performance seems unlikely on the basis of our previous runway experience. One major difference between this and other studies which have found shift effects of the opposite sort (e.g., Collier & Marx, 1959; Crespi, 1944; Zeaman, 1949) is that these *Ss* had run 64 trials before the shift. Spence (1956) has shown a similar effect for his shift-up group after 48 trials. Unpublished studies from our laboratory show that the same variable reduces shift effects in the Skinner box. A second difference between this and other shift studies is that these *Ss* were run two trials a day on alternate days. The failure to demonstrate over- or undershooting may be the result of the second trial which offers *Ss* two samples of the change in reinforcement "close" together in time, differing in this respect from the studies of Crespi (1944), Zeaman (1949), and Spence (1956). That this variable has some effect can be seen in the periodicity associated with the two trials.

A surprising finding in the present study is that the course of change differed radically for the shift-up and shift-down groups. The shift-up was accomplished rapidly, the shift-down gradually and with a periodicity (in starting times) suggesting that the rat starts each day at a level commensurate with the previous reinforcement.

The data on the consummatory response did not parallel the running and starting data. The rats shifted from 32% to 4% showed increased variability and a dramatic drop to a level below the prior 4% *Ss* with recovery over trials to the 4% level. The *Ss* shifted to 11.3% became much more variable but showed an average level of responding not much different from their previous level. Both groups of *Ss* shifted to

32% showed a gradual rise to a level of consummatory activity above the prior 32%.

On the basis of these findings it may be concluded that variations in the total amount of consummatory activity are not necessary conditions for reinforcement. Neither do they necessarily parallel variations in the instrumental activity.

Summary

Two groups of rats were run in a 16-ft. runway to a drinking tube. One group received 4% sucrose and the other 32%. Both groups were allowed a fixed time on the drinking tube and the number of licks, amount consumed, and the times to traverse successive segments of the runway were measured. Following training, a delay, just preceding the endbox, was introduced. Following retraining after elimination of the delay, a shift in the value of reinforcement was introduced.

Variation in amount of consummatory activity over a fixed interval of time was not consistently associated with variation in running. The high concentration group always ran faster but drank slower (less) at first and faster (more) at the end of training. The introduction of delay did not result in variations in consummatory activity commensurate with those produced in running. The shift in reinforcement magnitude appeared to be mainly disruptive of consummatory activity, whether the shift was up or down. On the other hand, the shift-up groups showed a rapid increase in running speed to a level below that expected from previous asymptotic performance, while the shift-down groups showed a gradual drop to a level above that expected. The gradient of running to the endbox was such that under all conditions the fastest running occurred in the early portion of the runway. The steepest gradient occurred with the greatest magnitudes of reinforcement. The steepness increased over trials.

It was concluded that variation in total consummatory activity is not a necessary condition for variation in effectiveness of food reinforcement.

References

Collier, G., & Marx, M. H. Frustration and performance. *Amer. Psychologist*, 1957, **12**, 458. (Abstract)

Collier, G., & Marx, M. H. Changes in performance as a function of shifts in the magnitude of reinforcement. *J. exp. Psychol.*, 1959, **57**, 305–309.

Collier, G., & Myers, L. The loci of reinforcement. *J. exp. Psychol.*, 1961, **61**, 57–66.

Crespi, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

Estes, W. K. Analysis of the relation between amount of reinforcement and rate of conditioning in a bar-pressing situation. *Amer. Psychologist*, 1954, **9**, 361. (Abstract)

Goodrich, K. P. Running speed and drinking rate as functions of sucrose concentration and amount of consummatory activity. *J. comp. physiol. Psychol.*, 1960, **53**, 245–250.

Knarr, F. A., & Collier, G. Taste and consummatory activity in amount of reinforcement effects. Paper read at Midwestern Psychological Association, St. Louis, April 1960.

Premack, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.*, 1959, **66**, 219–233.

Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

Swindells, J. F., Snyder, C. F., Hardy, R. C., & Golden, P. E. *Viscosities of sucrose solutions at various temperatures: Tables of recalculated values*. (Supplement to National Bureau of Standards Circular 440) Washington, D. C.: United States Department of Commerce, 1958.

Wist, E. R., & Collier, G. Delay and amount of reinforcement in the runway. Paper read at Midwestern Psychological Association, St. Louis, April 1960.

Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received August 22, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 496–502

FREE VERSUS SERIAL RECALL

NANCY C. WAUGH[1]

Harvard University

It has recently been shown (Waugh, 1961) that the serial memorization of digits proceeds at a constant rate, trial by trial. On each successive trial with a long list, the experienced *S* learns a new group of items farther down the list. Since the size of this group is constant, and since he seldom forgets an item once he has learned it, the total number of items that he can recall increases linearly with the number of trials. The curve for serial learning can in this case be considered to represent the number of trials multiplied by *S*'s span of immediate memory, up to the point where he masters the list.

With free recall, on the other hand, there is evidence that the learning curve is not linear but negatively accelerated (Bruner, Miller, & Zimmerman, 1955; Bush & Mosteller, 1955; Miller & McGill, 1952). Why this discrepancy? One obvious difference between the free and the serial recall experiment is the order in which *S* sees the items that he is to learn. Thus, in Bruner, Miller, and Zimmerman's (1955) study, the order of presentation was changed before every trial, while in tests of serial learning it is, of course, kept constant. Does the difference between free and serial recall depend on the order in which *S* sees the items on successive trials? Or does it depend instead on the order in which he is instructed to recall them? These alternatives were tested in the present study by comparing the learning curves obtained under the following three conditions: serial recall (S), free recall with the items to be recalled scrambled before every trial (F_v), and free recall with the order of the items kept constant (F_c). The items, monosyllabic English words, were presented and recited for a constant number of trials. The same group of *S*s served under each of the conditions, in order that the groups be directly comparable.

METHOD

The *S*s were nine college students, five men and four women, who were paid for their services. The material that they were to learn consisted of 18 series of words, each made up of 48 randomly ordered monosyllables. Half of these words were drawn from lists compiled by Egan (1948); the other half were drawn at random from a list compiled by Moser, Dreher, and Oyer (1957). No item ever appeared in more than one series.

For each of the three experimental conditions, a total of six series were exposed, each one for six trials. There should thus have been a total of 54 lists recited, six by each of 9 *S*s, under every condition. These numbers were actually 53, 53, and 52 for Cond. S, F_c, and F_v, respectively, due to absences by 2 *S*s.

The words were flashed onto a screen one at a time by means of a slide projector and an associated timing device. Each was exposed for 1.5 sec. and was followed by a blank interval of .5 sec. The total time between the onset of one word and the next was therefore 2 sec. After the series of 48 words had been shown, there was a 3-min. interval during which *S*s were to write down the words that they had just seen.

All *S*s learned the same lists in the same order. In every session, they recited two lists, each under a different experimental condition. The 18 lists therefore required nine sessions. The conditions were presented in the following order: F_c, F_v; S, F_c; F_v, S; F_c, S; F_v, F_c; S, F_v; F_c, F_v; S, F_c; and F_v, S.

[1] This work was carried out while the author was at the Lincoln Laboratory of the Massachusetts Institute of Technology.

Thus each was paired with each of the others three times; each appeared first within a session three times; and each appeared three times within each half of the experiment. There was a 15-min. interval between lists within each session.

The *S*s were instructed, under Cond. F_c and F_r, to recall as many words as they could in any order that they chose. Under Cond. S, they were instructed to recall the words in their proper serial positions (guessing, if necessary) and to write down early words before late ones. They were not specifically enjoined to attend to each successive word as it was exposed.

Each *S* was given two 6-page answer booklets, with 48 numbered blanks on each page, along with a copy of the general instructions, at the beginning of every session. The *S*s were forbidden to refer back to a given answer sheet after the next trial had begun, and were not allowed to write words down except during the interval between trials. The paper on which they wrote was quite thick, so that writing on one page did not produce legible impressions on the page underneath.

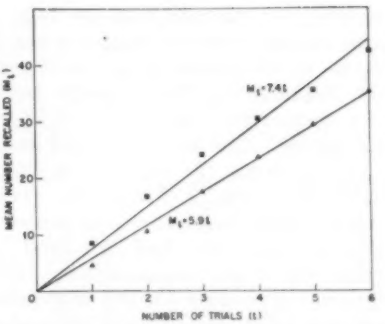

FIG. 1. Mean number of words recalled as a function of the number of trials, under Cond. S (serial recall). (The triangles denote the number recalled in correct serial positions, while the squares denote the total number recalled. The value shown for the total number recalled on Trial 6 is actually a median, rather than a mean: the distribution of scores on that trial was highly asymmetrical.)

RESULTS

Serial recall.—The data obtained under Cond. S (serial recall) are summarized in Fig. 1. Here the lower set of points represents the mean number of words, out of a list of 48, that were recalled in their correct serial positions on each trial. The upper points represent the mean total number of words recalled per trial, both in correct and in incorrect serial positions. The smooth functions in Fig. 1 are best-fit straight lines through the origin, calculated for each set of data by a method of least squares. These lines clearly provide a good description of the data.

Note that the number of words recalled in correct serial positions represents a constant proportion of the total number recalled. This total is thus the sum of two variables—the number in correct and the number in wrong serial positions—each of which is a linear function of the number of trials. It is reasonable to suppose that improperly placed words were actually remembered in the correct order relative to each other, rather than haphazardly, and were simply displaced on the answer sheet. No attempt has been made to test this hypothesis.

Table 1 shows the results of a detailed analysis of these data that was performed with the aid of an IBM 709 data processing machine. The input to this computer consisted of the number of each trial on which each *S* recalled each word in the proper serial position. Since there were six trials, there were 64 possible patterns of recall for a given word, ranging from recall on no trial to recall on all six. The data were reduced to a table showing the proportion of words associated with each of these 64 patterns, for all *S*s and all lists. These proportions were then combined to give the values shown in Table 1. Here each entry represents the proportion of words

recalled for the first time on Trial j and recalled also on Trial t, t being equal to or greater than j. Thus, for instance, the first column gives the proportion of words recalled for the first time on Trial 1 and then (reading down the column) recalled again on Trials 2 through 6. The right margin shows the total proportion of words recalled on Trial t, while the main diagonal gives the proportion recalled for the first time on each trial. The number of cases on which these proportions are based (2544) is given by the number of subject lists (53) multiplied by the number of words per list (48).

The purpose of this analysis was to test the hypothesis that the curve of serial learning is linear because (*a*) words once learned are never forgotten and (*b*) a constant number of new words are learned on every trial, so that the total number of words recalled after a given trial is simply this constant number multiplied by the number of the trial. According to this hypothesis, therefore, the entries in Table 1 should all be identical, while the marginal totals should increase linearly. Obviously the hypothesis is somewhat too simple.

It appears, for one thing, that people sometimes forget a word after they have recalled it once. They are likeliest to do so on the very next trial, and are less likely to do so on later trials. This effect is evident in the columns of Table 1. Here we see, for example, that, on the average, about six words (12.4% of 48) are recalled for the first time on Trial 2. On the next trial, only about five of these six words are remembered. Then the proportion recalled increases gradually, until the original six words are again almost always recalled on Trial 6. This increase is consistent with the well-known principle that the more times a word has been recalled already, the likelier it is to be recalled again.

TABLE 1

PROPORTIONS OF WORDS RECALLED ON EACH TRIAL (t), ANALYZED ACCORDING TO THE TRIAL ON WHICH THEY WERE FIRST RECALLED (j): COND. S (N = 2544)

t	j						Total
	1	2	3	4	5	6	
1	.105						.105
2	.096	.124					.221
3	.099	.107	.167				.373
4	.103	.116	.147	.143			.509
5	.101	.120	.147	.122	.127		.617
6	.103	.123	.157	.132	.117	.103	.735

It is also apparent that the number of words recalled for the first time on each trial is not perfectly constant. If we follow the main diagonal of Table 1, we see instead that this number, expressed as a proportion, increases up through Trial 3. It then decreases, so that by Trial 6 the proportion of words recalled for the first time is almost exactly what it was on the first. The departure of these six proportions from equality is highly reliable ($\chi^2 = 58$; $P < .001$). The direction of this departure, furthermore, is contrary to what we might have expected. Retaining words already learned does not seem to interfere with learning new ones. Rather, it appears that mere exposure to the individual items in a list may facilitate the forming of associations between them, an effect that has been reported by Hovland and Kurtz (1952) and by Noble (1955). Note that the two effects evident in Table 1—the forgetting of old items and the recalling of more new ones on intermediate trials—are complementary, so that the total proportion of items recalled does increase linearly with the number of trials.

Free recall.—Free recall presents a very different picture. Figure 2 shows

the mean number of words, estimated across *S*s and lists, that were recalled under Cond. F_c (constant order of presentation) and F_v (variable order). These values, denoted by the circles and the triangles for Cond. F_c and F_v, respectively, have been fitted with a simple negative exponential curve. A striking feature of Fig. 2 is the extreme similarity of the two sets of data. None of the small differences between corresponding pairs of points are statistically reliable. The inference from these data is either that scrambling the words in a list before every trial does not retard learning, or that keeping them in the same order does not facilitate it, when learning is tested by free recall. Repeated contiguity between the items to be learned is in this instance an unimportant variable.

Let us now compare Fig. 2 with Fig. 1. It is evident that, on early trials, more words are recalled, on the average, under Cond. F_c and F_v than under Cond. S—a fact consistent with Deese's (1957) data on the oral recall of spoken words after one trial. Note, however, that by Trial 6 the total number recalled is greater when

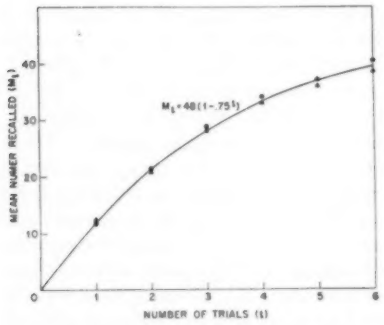

FIG. 2. Mean number of words recalled as a function of the number of trials, under Cond. F_c and F_v (free recall). (The circles represent Cond. F_c—words presented in a constant order; the triangles represent Cond. F_v—words presented in a variable order.)

TABLE 2

PROPORTIONS OF WORDS RECALLED ON EACH TRIAL (t), ANALYZED ACCORDING TO THE TRIAL ON WHICH THEY WERE FIRST RECALLED (j): COND. F_c ($N = 2544$)

t	j						Total
	1	2	3	4	5	6	
1	.237						.237
2	.208	.244					.452
3	.205	.207	.185				.597
4	.214	.204	.151	.136			.705
5	.208	.207	.154	.112	.089		.770
6	.221	.219	.163	.116	.074	.045	.838

recall is serial than when it is free. An appreciable number of the lists were, as a matter of fact, recited perfectly on this trial under Cond. S, whereas almost none of the lists were ever mastered under Cond. F_c and F_v. Serial recitation, therefore, was eventually the more effective method of study in this experiment.

The data on free recall are further analyzed in Tables 2 and 3, which are analogous to Table 1 and were constructed in the same way. Here again each entry represents the proportion of items recalled for the first time on Trial j and also recalled on Trial t, for $t \geq j$. A comparison of Table 2 with Table 3 reveals differences of very small magnitude. Neither the rate at which old items are forgotten nor the rate at which new ones are acquired appears to depend in any way on whether the list is presented in a constant or in a different order on every trial. Tables 2 and 3 show also that the proportion of new words recalled on each trial decreases with the number of the trial, from an average of .249 on Trial 1 to an average of .048 on Trial 6.

If we now compare Tables 2 and 3 with Table 1, we see that relatively more words seem to be forgotten when recall is free than when it is serial. (Thus, for example, about

TABLE 3

PROPORTIONS OF WORDS RECALLED ON EACH TRIAL (t), ANALYZED ACCORDING TO THE TRIAL ON WHICH THEY WERE FIRST RECALLED (j): COND. F_v ($N=2496$)

t	j						Total
	1	2	3	4	5	6	
1	.261						.261
2	.215	.233					.448
3	.218	.181	.188				.587
4	.229	.190	.143	.126			.688
5	.233	.195	.159	.093	.077		.757
6	.229	.205	.162	.098	.057	.051	.802

10% of the words recalled on Trial 1 under Cond. S are forgotten on Trial 2, while under Cond. F_e and F_v an average of about 16% of the words recalled on Trial 1 are forgotten on the next trial.) This inference was tested by the following procedure. The number of words recalled for the first time on Trial j and either (*a*) recalled again on Trial $j+1$ or (*b*) forgotten on that trial, under (*a*) Cond. S or (*b*) Cond. F_e and F_v, were combined in a 2×2 contingency table. Five such tables were constructed, one for every j from 1 to 5. The value of chi square then calculated for each table was in every case statistically significant ($\chi^2 = 7.45$, $P < .01$; $\chi^2 = 4$, $P < .05$; $\chi^2 = 16.63$, $P < .001$; $\chi^2 = 7.01$, $P < .01$; and $\chi^2 = 24.49$, $P < .001$, for $j = 1$, 2, 3, 4, and 5, respectively).

Both of the effects evident in Tables 2 and 3—the forgetting of recently learned words, the acquiring of fewer new ones on later trials—influence the total number of words recalled on each successive trial. The overall rate at which the list is learned thus decreases with the number of trials. The result is a negatively accelerated learning curve, in contrast to the linear curve that results when recall is serial. The difference between these curves depends not on the order in which the items are presented but rather on the order in which the learner must recall them.

The effect of word frequency.—Since the various lists used in this experiment were made up of words selected at random, it was assumed that each would contain both common and unusual words in about the same proportions. The effects of word frequency on speed of learning should thereby have been spread uniformly over the three experimental conditions. One may nevertheless wonder about the extent to which *S*'s previous familiarity with a word did affect the likelihood of his recalling it in this experiment.

In order to answer this question, the 864 words used in this study were sorted according to their frequency of usage in printed English, as reported by Thorndike and Lorge (1944). A count was then made of the number of trials elapsing before each word was recalled for the first time by each *S*. Thus, a word first recalled on Trial 1 by 1 *S*, and on Trial 6 by another *S*, was assigned values of 0 and 5, respectively. A word that was never recalled by a given *S* was assigned a value of 6. These values were averaged across *S*s and across words with Thorndike-Lorge counts of 0–9, 10–19, 20–29, 30–39, 40–49, 50–99, and 100+ occurrences per million words. There was no evidence that mean number of trials to first recall, as calculated by the above procedure, depended in any way on Thorndike-Lorge frequency, according to Edgington's (1961) test of differences in ordered series ($P = .56$, for four runs and seven observations).

Several earlier investigators have studied ease of learning as a function of word frequency. They have variously reported a positive monotonic relation (Hall, 1954), a U-shaped

function (Nachmias, 1958; Peters, 1936), and no relation (Underwood & Postman, 1960). The results of the present study agree with those of the latter. The procedures used in these various studies differ so widely, however, that disparate results are not necessarily contradictory.

Discussion

The data observed here on the serial learning of words are consistent with those reported previously on the serial learning of digits (Waugh, 1961): the number of items recalled in correct serial positions increases linearly with the number of trials. While the number of new words learned on each trial is not perfectly constant, but instead increases somewhat as *S* becomes familiar with the list, and while some forgetting does take place, the linear learning curve might nevertheless be said to represent a cumulation of memory spans. When recall is serial, the *S*'s usual strategy is to overlearn successive groups of words, starting at the beginning of the list, and to ignore later items while doing so. Thus 1 *S*, for instance, reported that he would "grab a mindful" of new words on every trial, rehearse this group, and then attach it onto the portion of the list that he already knew. The question that remains to be answered is just why this "mindful" that is assimilated on each successive trial should be limited to five or six words.

It may be an important fact that serial learning was here tested not by the usual method of anticipation but rather by what might be termed a modified method of complete presentation (McGeoch, 1942, p. 10). The *S* was not required to respond to each word as it came up but was instead free to distribute his attention as he pleased throughout the exposure trials. He could thus either ignore or overlearn whichever items he chose. It is possible, therefore, that the linear learning curve is an effect of the testing procedure used here. If a constant number of items were attended to and thus learned completely and discontinuously on each trial, while others remained unnoticed (and thus unlearned), the result would be a linear learning curve. In this connection, it is interesting to note that several recent studies which are said to demonstrate the all-or-none learning of paired associates have employed an exactly analogous testing procedure (Clark, Lansford, & Dallenbach, 1960; Estes, Hopkins, & Crothers, 1960; Rock, 1957). Now a method of anticipation would presumably require *S* to shift his attention to each successive item in a list, regardless of whether he had thoroughly committed the previous one to memory. Would either serial or paired-associate learning still appear to be discontinuous when tes ed by this method?

The major reason, of course, for not using a method of anticipation in the present investigation was to make the conditions of serial recall comparable to those of free recall. Obviously, it would not make sense to test free recall by a method of anticipation. The present data show that when all of the experimental conditions except the instructions to *S* are equivalent, free recall differs markedly from serial. In the former case, the learning curve is negatively exponential; in the latter, it is linear.

The present investigation has shown also that it does not matter whether a series is reordered before every trial or whether it is presented in a constant order, when recall is free. This surprising outcome implies that repeated contiguity between words cannot be an important principle of verbal learning when *S* is not required to associate contiguous words. Contiguity may perhaps be necessary or even sufficient for the forming of associations. It does not seem to be necessary, however, for the maintaining of associations. A word is recalled just as easily when it appears in a new context as when it reappears in the same context on every trial. Of course, it may be that multiple associations accrue to words that are scrambled before every exposure. If so, the present findings argue that these additional

associations neither facilitate nor inhibit recall.

An alternative way to interpret these findings is to assume that no associations are formed between any of the items in a series when recall is free, so that contiguity is in this case a completely irrelevant variable. The persistent occurrence of clusters of words on successive recall trials under Cond. F_V, evident on visual inspection of the *Ss*' answer sheets, renders this assumption an unlikely one.

Summary

The present study compares the free and serial recall of lists of monosyllabic English words over the course of six trials. In agreement with earlier findings, the learning curve for serial recall is linear, while the curve for free recall is negatively accelerated. The difference between these functions results from the fact that fewer new words are learned on each successive trial when recall is free, while an approximately constant number are learned when recall is serial. Relatively more words are forgotten on each trial, furthermore, when recall is free than when it is serial. Ease of learning, as denoted by the number of the trial on which a particular word was first recalled, is independent of this word's Thorndike-Lorge frequency under both conditions of recall.

Free recall was not found to depend in any way on whether the words to be learned were scrambled before every trial or were presented in the same order on all trials. The difference between free and serial learning must therefore result from the order in which the *S* is required to recite these words, rather than on the order in which he sees them. Since presenting a list in the same order on every trial does not facilitate its free recall, repeated contiguity between items to be learned does not appear in this case to be an important variable. It is concluded that serial learning represents a cumulation of memory spans. Free recall, on the other hand, is not so easily interpreted.

References

Bruner, J. S., Miller, G. A., & Zimmerman, C. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49**, 187–192.

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Clark, L. L., Lansford, T. G., & Dallenbach, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73**, 22–40.

Deese, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577–582.

Edgington, E. S. Probability table for number of runs of signs of first differences in ordered series, *J. Amer. Statist. Ass.*, 1961, **56**, 156–159.

Egan, J. P. Articulation testing methods. *Laryngoscope*, 1948, **58**, 955–991.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

Hall, J. F. Learning as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.

Hovland, C. I., & Kurtz, K. Experimental studies in rote-learning theory: X. Prelearning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.

Miller, G. A., & McGill, W. J. A statistical description of verbal learning. *Psychometrika*, 1952, **17**, 369–396.

Moser, H., Dreher, J. J., & Oyer, H. J. *One-syllable words.* Columbus: Ohio State University Research Foundation, 1957.

Nachmias, J. The effect of stimulus-heterogeneity on free recall. *Amer. J. Psychol.*, 1958, **71**, 578–582.

Noble, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

Peters, H. N. The relationship between familiarity of words and their memory value. *Amer. J. Psychol.*, 1936, **48**, 572–584.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

Waugh, N. C. Length of series and the learning curve. *Amer. J. Psychol.*, 1961, in press.

(Received August 22, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 503–509

THE EFFECTS OF AMOUNT OF REWARD, REQUIREMENT, AND SEVERAL RELATED PROBABILITIES ON HUMAN PERFORMANCE [1]

SHEILA G. ZIPF

University of California, Berkeley

The present study is an investigation of change in human performance as a result of variations in motivating conditions. Three of the four variables whose arrangement constituted the motivation conditions have been shown to be positively related to performance in animals. They are: amount of reward (e.g., Crespi, 1944; Hulse, 1958), probability of reward for a given performance (Hulse, 1958; Lewis & Cotton, 1957), and amount of work required in order to obtain the reward (Crutchfield, 1939). The fourth variable, probability of success, has been of great interest in recent human research but has been used to predict choice rather than performance.

A number of theorists have suggested that performance is a result of some combination, usually multiplicative, of amount or value of the reward and probability or expectancy of reward (Atkinson, 1957; Edwards, 1955; Lewin, 1938; Rotter, 1954; Tolman, 1955). Feather (1959a) reviewed these positions in a recent article. It is frequently overlooked, however, that the probability of reward (PrR) may be independent of the probability that the organism will be successful at the required task (PrS). In general, PrR is fixed by the situation while PrS is a result of the organism's ability at the task. Both of these probabilities are sufficiently meaningful for human *Ss* to report what they think they are and for each of these reports to correlate significantly and independently with change in performance (Zipf, 1960). As to the combination of these two probabilities, it is reasonable to expect that they multiply together. This joint probability may be thought of as the probability of obtaining the reward if the organism attempts the task.

Although usually held constant, it is clear that within limits, the more work that is required to obtain a goal, the more is done. This dimension might well interact with amount of reward, PrR and PrS. The experiment was, therefore, designed to observe the effects and interactions of these four variables.

METHOD

Design.—A $3 \times 2 \times 2 \times 2$ factorial design was used. PrS was given three levels (5%, 30%, 95%) because of some indication from earlier data that a PrS of around 30% produced less change in performance than lower or higher probabilities. The rewards offered were either 10¢ or $1. The PrR was set at 50% or 90%. Subjects were required to go 1.1 or 2 times as fast as they had been going in order to obtain the reward. A positive relationship to change in performance was expected for each variable.

Because this design diverges considerably from most studies of performance, the differences require clarification. A primary dif-

[1] This investigation was supported by a postdoctoral fellowship (MF-9450) from the National Institute of Mental Health, United States Public Health Service and was conducted at the University of Rochester. The writer wishes to express appreciation to Vincent Nowlis, who sponsored and encouraged this research, and to the Department of Psychology, University of Rochester, for the facilities provided.

ference is that the amounts, probabilities, and requirement for reward were not learned over time through experience but rather were given directly by instructions. This method is often used in human studies of choice.

Speed on the task was the performance to be considered and the reward was offered for a particular speed. This differs from the usual animal study in which the reward is given for getting to the end of the runway or for making the correct choice, and the speed which is observed does not affect whether the animal obtains the reward or not. Therefore, the relationship of motivation to speed of performance is more direct in the present situation.

The effect of motivation on performance vs. learning was not of concern here, and *S*s were not run to an asymptote prior to changing the motivational conditions. Since opportunity to learn is the same in all conditions, it does not confound the results. It will lead to inaccuracy if different *S*s are more or less able to change their speed because they are at different stages of learning. The following procedures were designed to reduce effects of learning and of individual differences. A simple task was chosen. The *S*s were told how to perform the task and asked to do it that way only. Everyone had 12 min. of practice prior to the test trials. Since the details of amount of reward, probabilities, and requirement were given in instructions, they did not have to be learned. Furthermore, absolute scores on the task are not used. Rather, each *S*'s score is his speed following the change in motivating conditions divided by his speed prior to that. This index is intended to control for differences in ability and initial motivation.

Task.—A card sorting task provided the measure of performance. The *S*s were given a deck of IBM cards having from 16 to 25 holes each, and a rack with 10 slots labeled 16, 17, . . . 25. The deck of cards was in a box which had a marker on the back that could be used to indicate how many cards had been removed. The task was to count the number of holes in each card and place it in the appropriate slot in the rack. The *S*s were asked to take about 10 cards in hand at a time, place one card face down on the table, and count the holes by twos. They were told that it was important to the experiment that it be done this way. Each frequency of holes was punched in three different patterns so that the frequency could not be recognized from the pattern. The decks which *S*s received were thoroughly mixed for frequency and pattern. The task had been used in several earlier experiments (French, Morrison, & Levinger, 1960; Zipf, 1960).

Procedure.—Subjects were assigned to sessions in groups of size 2 to 12. The task was explained and they were told that they might earn some money later. They sat in separate booths. There were three work periods of 6 min. each, with breaks between each minute and a longer break between work periods. During these breaks, different colored cards were placed in each slot so that work could later be identified as to the minute in which it was done. The markers were set to zero before each 6-min. work period. With them, *S* was able to see approximately how many cards he had sorted during the work period.

Before *S*s started to work, they were told to sort at a comfortable speed and to be accurate. Before the second work period, they were given an instruction sheet telling them to be accurate and sort at a speed that was comfortable, but reasonably fast. After the second work period, *E* went around to each booth, looked at the marker behind the deck of cards, and gave each *S* an instruction sheet (see below). The *S*s were randomly assigned to the 24 conditions described by these instructions. A brief questionnaire followed the instruction sheet, and another was filled out following the third work period.

Instruction sheet.—The following is a copy of the instruction sheet used before the third work period. All of the alternatives are indicated, but only one appeared in each position on the actual sheets. The italicized portions were filled in by *E* while standing beside *S*.

> Last time you sorted about *X* cards. In the next work period, which will also be six minutes long, I would like you to try to sort one and one tenth (two) times as many as that correctly. That is, I would like you to try to sort *N* cards correctly next time.
>
> If you had about average accuracy in the last work period, from what I know about how others have done with this amount of practice and under these conditions, your chances of being able to sort the requested number of cards next time are *very low* (*low*) (*high*). That is, the probability that you will be able to count *N* cards correctly next time is about *5* (*30*) (*95*) %.
>
> If you do sort that many cards correctly next time, you may receive a reward of 10¢ ($1) before you leave. If you do sort the requested number of cards cor-

rectly, I will give you a folded slip of paper with a number from 1 to 10 on it. You may have five (nine) guesses at the number written on it. If you guess it, you will receive the dime (dollar). This means that you have a 50 (90)% chance of getting the dime (dollar) if you count the requested number of cards correctly. Only cards that are sorted correctly will be counted in figuring your total.

After you have answered the questions on the next page: keep this sheet where you can see it so you will remember what it says. Then, adjust the marker paper behind the cards so that the top line on the paper is even with the top of your stack of cards. This will show how many cards you are counting as you go along.

Subjects.—The *S*s were 289 men and women volunteers from freshman English and introductory psychology classes at the University of Rochester. Men and women were always in separate sessions. They were told that the purpose of the experiment was to study the effects of different practice and working conditions on routine tasks. Five *S*s had to be dropped when their performance data were lost through some error by *E* or *S*. To equalize cell *N*s, only 10 *S*s per cell could be used. The first six males and four females who had been obtained in each condition, 240 *S*s in all, were used in analysis of the data.

Results

An analysis of variance was done on the ratio: total number of cards *S* sorted in the third work period divided by total number of cards *S* sorted in the second work period. The only significant effect was required speed (Req). Since experimental instructions were quite complicated, it seemed likely that the variables would have their effect most clearly immediately after the instructions were given. The ratio: total number of cards sorted in the first minute of the third work period divided by total number of cards sorted in the first minute of the second work period ($t_{3,1}/t_{2,1}$) was computed. The abbreviation $t_{i,j}$ means the total number of cards sorted in the jth minute of the ith work period. These scores were transformed to log $(1+X)$ which corrected a significant heterogeneity of variance. The results of analysis of variance on these transformed scores are given in Table 1. Probability of reward and Req each have significant effects. Probability of success is also significant, but the effect is opposite to what was expected. Amount of reward is not significant, nor are any of the interactions. One second-order interaction which was just significant using the untransformed scores is just below significance with the transformed scores. This interaction appears to be due to the irregularity of effects in Low Reward as compared to High Reward conditions.

Since there are no interactions, the variables appear to be additive rather than multiplicative. At least one animal study (Hulse, 1958) failed to obtain an interaction of two of these variables: amount of reward and PrR. However, it remains possible that better controlled research with a wider range of the independent vari-

TABLE 1

Analysis of Variance of Transformed Performance Scores: Log $(t_{3,1}/t_{2,1}+1)$

Source	*df*	*MS*	*F*
Amount of reward	1	000	—
Probability of reward (B)	1	.026	8.67**
Probability of success (C)	2	.013	4.33*
Requirement (D)	1	.019	6.33*
A × B	1	.006	2.00
A × C	2	.002	.67
A × D	1	.002	.67
B × C	2	000	—
B × D	1	.002	.67
C × D	2	.001	.33
A × B × C	2	000	—
A × B × D	1	000	—
A × C × D	2	.009	3.00
B × C × D	2	000	—
A × B × C × D	2	.004	1.33
Error	216	.003	

* $P < .05$.
** $P < .005$.

ables will support the multiplicative hypothesis.

Amount of reward.—As is often the case with human *S*s, need or desire for the incentive was not controlled in this study. It is thus possible for variability in desire for the reward to obscure the effect of amount of reward. Using verbal reports, we can infer this from these data. Immediately after being informed of the reward conditions, *S*s were asked how much they wanted the reward. Answers were checked on a nine-point scale. Table 2 shows the distribution of high and low reward groups on this dimension broken as nearly as possible into thirds. Although the larger reward is generally more preferred, there are many who want 10¢ more than others want $1. Presumably the dime is valued as a symbol of success. Desire for the reward has a low but significant correlation with the transformed performance scores across all conditions ($r = .17$, $N = 214$, $P < .05$).

Another matter of interest concerning the amount of reward is the general regularity of the data in High and Low Reward. The other three variables are all closely connected to the monetary reward. PrR is clearly so. Req is the speed necessary to obtain the reward. PrS is the probability of reaching the Req in order to be eligible for the reward. Assuming that 10¢ is generally a less valuable reward than $1, even though it produces as much work by *S*, we might expect that relatively more nonmonetary factors are motivating *S* to work in Low Reward conditions than in High Reward. Need to achieve, desire to please the experimenter, and other personality factors may be involved. Regardless of what the other factors are, we would expect that PrR, Req, and PrS would be more important determiners of performance in High Reward than in Low. Table 3 presents the means of the transformed performance scores for the entire 24 cells. It is readily seen that within High Reward these three variables each have completely regular effects in line with the overall findings, while within Low Reward

TABLE 2

NUMBER OF *S*s IN HIGH AND LOW REWARD GROUPS WITH VARYING DESIRE TO OBTAIN THE REWARD

Amount of Reward	Desire for Reward			Total
	Low	Medium	High	
10¢	52	38	30	120
$1	17	33	70	120
Total	67	71	100	240

TABLE 3

MEANS OF TRANSFORMED PERFORMANCE SCORES IN EACH CONDITION

Req	PrR	10¢ Reward			$1 Reward		
		5% PrS	30% PrS	95% PrS	5% PrS	30% PrS	95% PrS
2.0	50%	.320	.320	.326	.367	.334	.320
	90%	.337	.369	.345	.373	.337	.327
1.1	50%	.328	.313	.280	.318	.316	.302
	90%	.364	.326	.330	.340	.336	.311

Note.—$N = 10$ per cell.

these variables have quite irregular effects.

Probability of reward.—In comparing the experimental groups on initial performance ($t_{2,1}$), it was found that those *S*s who later received the Low PrR instructions were significantly faster than those who later received the High PrR instructions. All *S*s had been treated alike before these instructions were given. They were randomly assigned to conditions and in each experimental session every *S* was assigned a different condition. It appears to be a chance effect, but, it makes interpretation of the PrR variable questionable. However, it is possible to draw a tentative conclusion. On the absolute test score ($t_{3,1}$) High and Low PrR do not differ. They have shifted from a pre-experimental difference to no difference, indicating that the High PrR *S*s increased more than the Low PrR *S*s, as expected.

As shown on Table 1, a significant effect is obtained for this treatment using the index $t_{3,1}/t_{2,1}$, which is a measure of relative change. However, this index takes advantage of the initially low score of the High PrR *S*s. The index is negatively correlated to $t_{2,1}$($r = -.41$, $N = 240$, $P = .01$). This is precisely the type of correction which the index is designed to make, giving more weight, for the same absolute increase by initially low scorers, than to initially high scorers. A more appropriate index would, of course, be independent of initial scores. Neither the assumptions made by the index nor the cause of the initial difference in the experimental groups are sufficiently understood to accept this result. Current research, utilizing more levels of PrR, indicates that change in performance is an increasing linear function of PrR.

Probability of success.—This variable did not show the relationship to performance that was expected. A trend analysis is not presented because of the unequal spacing of the levels. However, performance change is clearly a decreasing function of PrS. The trend appears to deviate slightly from linear, having the general form of a negative exponential function. (The means of transformed scores for 5%, 30%, and 95% are respectively: .341, .331, .318, $\sigma_{\bar{x}} = .006$.) The direction of this relationship is opposite to that found in earlier correlation studies. The lower the probability of success, the greater was the increase in speed. A measure of what *S* thought his probability of success was showed that the probability given in the instructions was generally accepted. However, the instructions seem merely to have indicated to *S* how hard he would have to work to achieve the goal. Right after reading the instructions about the conditions of reward, *S*s were asked how hard they would have to try in order to succeed. Using this measure as the dependent variable, an analysis of variance of the same form as that shown in Table 1 was done. Both Req and PrS had significant effects (Req, $df = 1$, $F = 121.14$, $P < .005$; PrS, $df = 2$, $F = 6.18$, $P < .005$). No other effects were significant. The larger the requirement and the less the purported chances of success, the harder *S*s felt they would have to try to succeed, and, indeed, the more they increased their speed.

The dependent variable.—It was decided prior to obtaining the data for this experiment that the dependent variable would be a ratio of speed on the test trial divided by speed on the trial prior to the test. This index had been used before (French et al.,

1960; Zipf, 1960) and is intended to control for the effects of individual differences in ability and motivation on the task prior to the reward instructions. It assumes that the faster *S* is going before motivating conditions are changed, the better able he will be to increase his speed subsequently. There is reason to accept this assumption, within limits. Several early studies on the effects of practice on performance of clerical tasks indicate that *S*s who are better at the beginning of practice improve more with practice than do the poorer *S*s (Anastasi, 1934; Thorndike, 1916).

It is instructive to compare the results obtained with this index of performance with several others that might have been used: absolute speed on the test trial $t_{3,1}$, and the difference between the test trial and the trial prior to change in motivation $t_{3,1}-t_{2,1}$. Using the same analysis as shown in Table 1 with these as dependent variables, there are no significant effects on $t_{3,1}$, and only Req is significant in its effect on $t_{3,1}-t_{2,1}$. Discounting the variable PrR for reasons given above, the ratio score obtains significance on two variables: the difference score on one variable and the absolute score on none. The ratio score would seem to be a more sensitive measure.

The use of the ratio is far easier than either matching groups or partialing out initial score with analysis of covariance, and possibly provides a better control. It is hoped that the index will lend itself to quantitative rather than merely directional predictions, which will apply to different tasks. Since the ratio would be seriously affected if *S*s were near an upper limit of speed, a measure of *S*'s upper limit of speed has been obtained in a more recent experiment. This measure will be used to clarify the type of index necessary to take account of differences in upper limit of performance as well as "operant level."

Discussion

The importance of this study is seen as providing a simple method of studying and scoring human performance both so that comparisons may be made with animal studies of motivation and performance and so that other factors which are more relevant to human behavior may be considered. The findings of animal research on the variables of the present study were mentioned above. The majority of studies of human behavior utilizing these variables have examined choice rather than performance. We might expect, however, that direction and intensity of behavior might have some commonality and comparison with these findings is of interest.

We have tentatively concluded that *S*s increase their speed more when the probability of reward is 90% than when it is 50%. Edwards (1953) has shown that when postive expected value of a bet is held constant, preferences have a nonlinear relation to probability of winning and that 50% is preferred over 90%. When the expected value of bets differ, *S*s do not bet so as to maximize expected value (Edwards, 1954). Probability preferences have been related to various personality characteristics (Atkinson, 1957; Scodel, Ratoosh, & Minas, 1959). If preferences of this nature are operating, they would tend to confound the effect of PrR on performance. However, none of the choice studies seem to have considered very simple decisions where expected value was clearly different as would be comparable to the values in this study, for example: 50% chance of getting $1 versus a 90% chance of getting $1. The writer has some preliminary data to indicate that in such simple decisions, almost every person chooses the higher probability. Furthermore, Feather (1959b) has shown that *S*s wish for less probable

goals but tend to choose more probable goals when the choice involves commitment to action. Another of his findings is compatible with a result of this study. In an achievement situation difficult tasks are preferred over easy ones while the reverse is true in a relaxed situation. The *Ss* in the present study who were told that they had a small chance of succeeding were probably in a more extreme achievement situation than were those who were told that they were very likely to succeed, and perhaps placed a greater value on doing well.

Explicit comparisons of how these factors affect choice and performance must await research in which both choice among tasks and performance on these tasks are observed. Such research will be in order when more is understood about factors affecting simple choice and performance situations.

Summary

Four variables were investigated for their effect on human performance. They were: amount of reward, probability of reward for successful performance, probability of success, and amount of increase in speed required. Probability of success and required increase in speed had significant effects on performance. No conclusion was possible for probability of reward with these data. Although amount of reward was not effective, a measure of desire to obtain the reward was significantly related to performance. A number of theories would suggest that the independent variables of this study would have multiplicative effects. However, no evidence of this was found.

Several different indices of performance were compared. A ratio of speed following the change in motivating conditions divided by speed before the change was shown to be more sensitive to the independent variables than either absolute speed or absolute change in speed.

References

Anastasi, A. Practice and variability. *Psychol. Monogr.*, 1934, **45**(5, Whole No. 204).

Atkinson, J. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

Crespi, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

Crutchfield, R. S. Determiners of energy expenditure in string pulling by the rat. *J. Psychol.*, 1939, **7**, 163–178.

Edwards, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

Edwards, W. Probability preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67.

Edwards, W. The prediction of decisions among bets. *J. exp. Psychol.*, 1955, **50**, 201–214.

Feather, N. T. Subjective probability and decision under uncertainty. *Psychol. Rev.*, 1959, **66**, 150–164. (a)

Feather, N. T. Success probability and choice behavior. *J. exp. Psychol.*, 1959, **58**, 257–266. (b)

French, J. R. P., Jr., Morrison, H. W., & Levinger, G. Coercive power and forces affecting conformity. *J. abnorm. soc. Psychol.*, 1960, **61**, 93–101.

Hulse, S. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

Lewin, K. The conceptual representation and the measurement of psychological forces. *Contr. psychol. Theor.*, 1938, **1**, No. 4.

Lewis, D., & Cotton, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. comp. physiol. Psychol.*, 1957, **50**, 189–194.

Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

Scodel, A., Ratoosh, P., & Minas, J. S. Some personality correlates of decision making under conditions of risk. *Behav. Sci.*, 1959, **4**, 19–28.

Thorndike, E. L. Notes on practice, improvability, and the curve of work. *Amer. J. Psychol.*, 1916, **27**, 550–565.

Tolman, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.

Zipf, S. G. Resistance and conformity under reward and punishment. *J. abnorm. soc. Psychol.*, 1960, **61**, 102–109.

(Received August 23, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 510–517

PERSEVERATION AS A FUNCTION OF DEGREE OF LEARNING AND PERCENTAGE OF REINFORCEMENT IN CARD SORTING [1]

ALBERT ERLEBACHER[2] AND E. JAMES ARCHER

University of Wisconsin

The purpose of this study was to investigate the effects of degree of learning and percentage of reinforcement, and especially their interaction, on perseverative errors in concept formation. Normally in the Wisconsin Card Sorting Test (WCST), *S* is required to sort a deck of cards according to the stimulus dimension *E* arbitrarily considers correct (Stage I). After *S* has achieved a criterion in Stage I, *E*, without informing *S*, changes the correct dimension (Stage II). If *S* erroneously continues to sort according to the formerly correct dimension in Stage II, the response is called a perseverative error.

Using the WCST, Grant and Berg (1948) and Grant and Cost (1954) had different groups of *S*s achieve different criteria before shifting to a new correct stimulus dimension. They found that there was an inverse relationship between the number of reinforced trials in Stage I and the number of perseverative errors in Stage II.[3] These findings seem to contradict the results of studies on simple learned responses in which a greater number of reinforcements usually leads to a persistence of the response. Grant and Cost (1954), however, believed that this inconsistency was only superficial. They suggested that *S* must learn to discriminate the change in the correct stimulus dimension and that this discrimination is facilitated by a large number of reinforcements.

This hypothesis is reminiscent of the hypothesis proposed by many researchers (e.g., Grant & Schipper, 1952; Humphreys, 1939) to account for increased resistance to extinction following partial reinforcement as opposed to 100% reinforcement in conditioning studies.

If Stage II is analogous to an extinction phase in simpler tasks we would predict that the number of perseverative errors should be inversely related to the percentage of reinforcement. Accordingly, three levels of percentage of reinforcement were investigated in the present study.

Furthermore, it seems very likely that the effect obtained under partial reinforcement is dependent on the number of trials during acquisition. It is not likely that there would be differential resistance to extinction under various percentages of reinforcement when only a few acquisition trials are given; if a larger number of trials were given however, there should be a differential resistance to extinction under various percentages of reinforcement. Perhaps this differential resistance to extinction even increases as a function of the number of trials. Indirect evidence for this hypothesis can be found in Grant and Berg (1948), and Grant and Cost (1954) who have shown that the

[1] This research was supported in part by NSF Grant G-1691 and while the first named author was a University Fellow.

[2] Now at Northwestern University.

[3] For similar results in animal learning studies see Reid (1953), Pubols (1956), and Capaldi and Stevenson (1957).

number of reinforcements influences the number of perseverative errors. These findings, however, followed 100% reinforcement, and it is entirely possible that these results would not obtain under other reinforcement schedules. In fact, if a low percentage of reinforcement is similar to extinction, then with a greater number of trials, there should be greater generalization to extinction and thus, the greater the number of trials given in a low reinforcement first sorting stage, the longer should be the perseveration. Thus it seemed advisable that in addition to percentage of reinforcement, the number of trials during the first sorting stage should also be varied. The most interesting part of the results should be the Percentage Reinforcement × Degree of Learning interaction.

To increase the generality of the results the experiment involved replications with WCST concept varied. The WCST allows cards to be sorted according to form, number, and color. Thus, *S* was started in Stage I on one of the three dimensions and then switched to one of the other two.

Method

Experimental design.—Three variables were presented according to a 3 × 3 × 6 orthogonal design. These variables were percentage of reinforcement (25, 50, 100%), degree of first stage learning (10, 20, 40 trials), and stimulus dimensions in the two stages (form-number, form-color, number-form, number-color, color-form, color-number). In addition, one reinforcement sequence was used for half the *S*s in each partial-reinforcement group while another sequence was used for the other half.

Subjects.—Two *S*s were run in each of the 72 treatment conditions under partial reinforcement and 4 *S*s were run in each of the 18 treatment conditions under 100% reinforcement. This makes a total of 216 *S*s. The *S*s were all students in undergraduate psychology courses at the University of Wisconsin and were assigned to the various treatment cells at random with the restriction that the experiment be completed once before it was replicated. A time limit of 35 min. was arbitrarily chosen for learning in Stage I. Any *S* who exceeded this time limit was dropped from the experiment. Twenty-five *S*s were dropped for this reason; 12 were to have sorted according to color and were in the 25% group, 5 were in color-50%, 2 in color-100%, and 6 in form-25%. Twenty-three *S*s were dropped for reasons not related to any experimental conditions, mainly because of apparatus failure.

Materials and apparatus.—The materials used were a modification of the WCST. The WCST consists of cards showing figures which vary in three stimulus dimensions: form (circles, triangles, crosses, and stars), number (1, 2, 3, and 4), and color (red, green, yellow, and blue). For example, one card might show four red crosses and another might have one green circle.

The WCST cards are normally sorted into four piles, the first labeled by one red triangle, the second by two green stars, the third by three yellow crosses, and the fourth by four blue circles. Thus, *S* can make one of four responses. He can sort according to form, number, or color, or he can make a response which is none of these and which is unanalyzable.

In order for each response to be unambiguously classified only 24 of the standard 64 cards can be used. For example, none may have a combination of red and one, or red and triangle, or one and triangle. Cards with combinations of levels for the other label cards were also eliminated. The set of 24 cards used in this experiment thus represented all possible combinations of the four forms, four numbers, and four colors which did not represent any of the above combinations.

The cards were not actually presented to *S*s, but were projected on the back of a milk-glass screen by a Dunning Animatic 16 mm. strip-film projector. The *S* sat in a 4 × 4 × 8 ft. booth directly in front of the screen, which was approximately 11½ in. wide and 7½ in. high. The pictures were approximately 7½ in. square.

The cards used were those of the "unsystematic" type (Grant & Curran, 1952). That is, the position of figures on a card was unrelated to the number of figures. The film strip was a closed loop and contained 192 pictures. Each combination of dimensions appeared eight times but unlike the standard WCST, four different configurations were used for each combination of dimensions.

A special procedure was followed in ordering the presentation of the cards. Essentially the entire sequence represented eight different orders of the 24 cards. Four sets of the 24 cards were made up, the sets differing only in card configuration. The 24 cards of one configuration set were randomly ordered. The cards of the other three sets were ordered systematically on the basis of this first order. The four sets of cards followed each other consecutively making up the first 96 cards. A different order for the second 96 cards was similarly obtained. A detailed account of the ordering procedure is presented by Erlebacher (1960).

Four response-label cards were placed under a clear plastic sheet across the center of *S*'s control box. The response cards from left to right were: one red triangle, two green stars, three yellow crosses, and four blue circles. At the rear center of the box was a white light; at the right rear of the box was a red light. In front of each response card was a button which *S* pressed to make his response. The white light was the reinforcement light and the red light indicated to *S* when the experiment was over. When a new picture appeared on the screen, *S* pressed one of the four buttons, taking as much time as he pleased. After he pressed the button, he waited until a new picture came into the window. If his response were reinforced, the white light came on at the moment of his pressing the button and it stayed on for approximately 2.25 sec. The next picture came into the screen approximately 2.60 sec. after the button was pressed.

The two partial reinforcement schedules were 25% and 50%. The percentage refers to the percentage of times *S* was told he was correct (the white light went on) *when he actually was correct*. Thus, whether any correct response was reinforced was contingent on *S*'s previous number of correct responses.

A schedule of reinforcement was made up which took into account 200 trials of correct responses on the part of *S*. These 200 trials were divided into blocks of 20 trials such that the percentage indicated was maintained in each block. The sequence of reinforcement-nonreinforcement was random with the restrictions that the first trial of each block be reinforced and that at the most, eight positive reinforcements could occur in a row. In order for the results of the experiment to be generalizable to various reinforcement sequences, half of the *S*s were started at the beginning of this schedule of 200 trials and the other half were started at the middle of the schedule. This difference in starting position represented the two reinforcement sequences mentioned earlier. The reinforcement schedule was repeated if *S* required more than 200 trials. All timing, reinforcement, etc., was controlled by automatic apparatus, consisting mostly of relay circuits, which was kept in a remote room away from *S*, so that *S* could not hear the apparatus in operation.

Procedure.—After *S* made either 10, 20, or 40 consecutively correct responses in Stage I the "correct" dimension for sorting was changed. Except for the change in reinforcement *S* was given no information about this change. He was now reinforced for every correct sorting on one of the other two stimulus dimensions and continued on this second stage until he had made 10 consecutively correct responses. If *S* did not reach this latter criterion at the end of 300 trials on the new dimension, the experiment was terminated.

Instructions.—The *S* was told that for each of the pictures appearing on the screen he was to press one of the four buttons in front of him and that only one of the four buttons would be correct. If he were correct, the white light might go on, but it would not necessarily do so. If he were incorrect, however, the white light would not go on. (It was necessary to give *S*s partial-reinforcement instructions, i.e., tell them that the white light would not always go on even if they were correct, because a preliminary pilot study indicated that most *S*s could not reach even a moderate criterion—especially in the 25% reinforcement group—if these instructions were not given.) All *S*s, including those of the 100% group, were given partial-reinforcement instructions. No mention was ever made of stimulus dimensions, and if *S* asked a question referring to such dimensions, he was not given a direct answer involving such a reference. No instructions about a change in the correct stimulus dimension were given.

Results

The responses during Stage I (learning) and Stage II (extinction) were analyzed separately. In Stage I, the responses were analyzed in three categories: number of correct responses, number of errors, and total number of responses. These measures do *not* include the criterional trials, since these were different for various

groups. In Stage II, the analyses were on number of perseverative errors, number of nonperseverative errors, and total number of responses. The criterional trials were included in Stage II measures. The first response of Stage II was considered a perseverative error since this would not differentially affect the scores of the various groups.

Stage I.—The total variance for Stage I was analyzed and showed that percentage of reinforcement was a significant variable affecting all three response measures ($P < .001$). Figure 1 presents the mean values for all three measures for all three reinforcement groups. As the percentage of reinforcement increases, all three scores decrease. It must be noted that the values for the 25% group are somewhat underestimated since the scores of those *S*s who did not reach the criterion (most of whom were in the 25% group) are not included.

The degree of learning was also a significant variable for total correct responses. The means were 20.0, 16.2, and 34.0 for the 10, 20, and 40-trial groups, respectively ($F = 4.83$, $df = 2/171$, $P < .01$).

Another significant term in the analysis was the Percentage Reinforcement × Dimension interaction for total responses ($F = 2.44$, $df = 4/171$, $P < .05$). This relationship is given in Fig. 2, which shows that the large factor contributing to this interaction is that sorting according to number was more difficult at 50% reinforcement than was sorting according to form, while on 25% reinforcement the order was reversed. Again it must be pointed out that the figure for total responses when sorting according to color is underestimated especially under 25% reinforcement, as is also the figure for form on 25% reinforcement.[4]

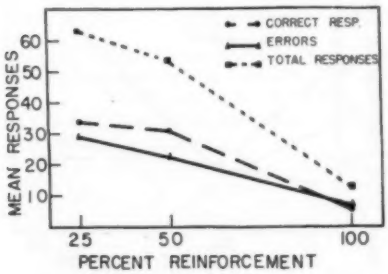

FIG. 1. Mean total responses, mean total correct responses, and mean total errors in Stage I as a function of the percentage of reinforcement.

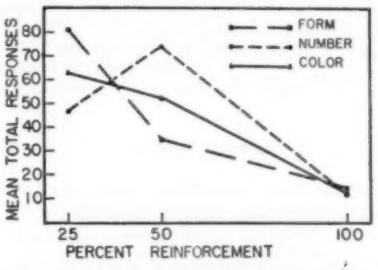

FIG. 2. Mean total responses in Stage I as a function of the percentage of reinforcement and the correct stimulus dimension.

Stage II.—Table 1 summarizes the analyses of variance of the three response measures obtained during Stage II.[5] The term *dimension* in

[4] Several significant interaction terms in the analysis involve sequence within percentage reinforcement which seems to indicate that the sequence used is effective according to what conditions it is used with. No estimation analysis is given for these interactions or for any terms in this paper involving sequences since sequences were not systematically varied and thus no conclusions could come from such an analysis. The reason for having introduced various sequences was only as a check for the generality of the results and the analyses of variance seem to indicate that any future studies on partial reinforcement, in this area, should involve several sequences of reinforcement.

[5] Since the experiment was terminated at the end of 300 responses, in order to obtain

TABLE 1

ANALYSES OF VARIANCE OF TOTAL RESPONSES, PERSEVERATIVE ERRORS, AND NONPERSEVERATIVE ERRORS IN STAGE II

Source of Variation	*df*	*F* Ratios		
		Total Responses	Perseverative Errors	Nonperseverative Errors
Percentage reinforcement (%)	2	1.90	2.90	.82
Degree of learning (DL)	2	1.98	3.49*	.64
Dimension (D)	5	1.87	.77	4.74**
Sequence/Percentage reinforcement (S/%)	2	.14	.33	1.50
% × DL	4	3.33*	3.44*	1.96
% × D	10	.66	.67	1.06
DL × D	10	1.53	1.51	1.61
DL × S/%	4	.20	.23	.33
D × S/%	10	1.10	.74	1.01
% × DL × D	20	.94	.84	.66
DL × D × S/%	20	1.20	1.02	.67
Error (*MS*)	126	(6986)	(5388)	(720)
Total	215			

* $P = .05$.
** $P = .01$.

Table 1 refers to a particular dimension sequence such as color-form, i.e., *S* had to sort according to color in Stage I and according to form in Stage II.

The analyses show that the degree of Stage I learning is a significant variable in determining the number of perseverative errors. The mean total perseverative errors were 49.9, 35.99, and 68.19 for the 10, 20, and 40-trial criterion groups, respectively.

The Percentage Reinforcement × Degree of Learning interaction is also a significant term in the analysis of variance for both total responses and total perseverative errors. This relationship is shown for perseverative errors in Fig. 3. It should be noted that the groups do not rank 100%, 50%, 25% going from fewest to most perseverative errors for *all* degrees of original learning. Such an ordering would have been expected if the normally found partial-reinforcement effect were so strong that it overshadowed any effect due to other

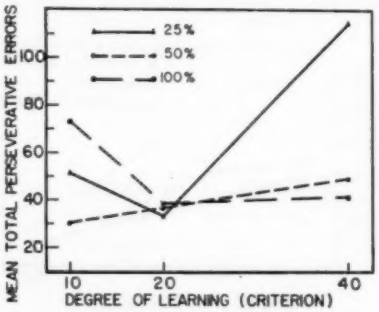

FIG. 3. Mean total perseverative errors in Stage II as a function of the percentage of reinforcement and the degree of learning in Stage I.

unbiased results for those *S*s who had not reached criterion, the scores were changed to percentage out of 300. This kind of a score normally requires an arc sin transformation, so a second analysis was performed on the transformed scores. The entire transformation consisted of dividing each score by 300, taking the square root of that result, and then finding the arc sin of that square root. This second analysis gave essentially the same results as are shown in Table 2. Since no different conclusions are reached when transformed scores are used, the results given below will be only for untransformed scores.

TABLE 2

NUMBER OF *Ss* MAKING 300 RESPONSES STAGE II

Degree of Learning (Trials)	% Reinforcement			Total
	25	50	100	
10	2	0	4	6
20	0	2	1	3
40	2	2	2	6
Total	4	4	7	15

variables. Also to be noted is the fact that the difference between the 25%, 50%, and 100% reinforcement groups does not increase as the number of criterional trials increases.

Fifteen *Ss* made 300 responses, at which time the experiment was terminated. Table 2 shows in which condition these *Ss* served.

The mean total perseverative errors were 66.3, 36.9, and 50.9 for the 25%, 50%, and 100% reinforcement groups. These differences were not significant.

The interaction for total responses is similar to that for perseverative errors. Since there is little difference in the shape of the curves for total responses and those shown in Fig. 3, separate figures are not presented here.

In addition to these variables, it is shown in the analyses of variance table that there is a significant effect due to the dimension sequence but only as far as nonperseverative errors are concerned. This relationship is shown in Table 3.

TABLE 3

MEAN TOTAL NONPERSEVERATIVE ERRORS AS A FUNCTION OF THE CORRECT STIMULUS DIMENSIONS IN BOTH STAGES

Correct Dimensions		Mean Total Nonperseverative Errors
Stage I	Stage II	
Form	Number	5.56
Form	Color	8.83
Number	Form	3.06
Number	Color	29.72
Color	Form	5.78
Color	Number	10.61

DISCUSSION

Stage I.—As was seen, performance in Stage I was affected by the percentage of reinforcement. This result was not unexpected. As the percentage of reinforcement increased, both the number of errors and number of correct responses decreased. Though it may seem reasonable that errors should decrease and therefore paradoxical that correct responses decrease also, the apparent inconsistency is due to requiring *S* to achieve a criterion.

It is also to be expected that the more stringent the criterion, the longer it takes to reach it. This relationship held for the highest criterion (40 consecutively correct responses). No difference, however, was seen between the two other criteria (10 or 20 consecutively correct responses).[6] The three criterion groups need not have differed in number of errors. Because the criterion was one of 10, 20, or 40 consecutively correct responses, one or two errors between the attainment of a lower criterion and the attainment of the final criterion could greatly increase the number of correct responses made, but have little effect on the error score. Such slight differences in number of errors could not show up reliably in this experiment because of the relatively large variability among *Ss*.

Stage II.—It was seen that the degree of learning in Stage I resulted in a differential number of perseverative errors in Stage II. Unlike previously cited studies, the groups with many overlearning trials (40 consecutively correct

[6] Although the number of correct responses required to reach criterion by the 20-trial criterion group was slightly less than by the 10-trial group, this difference can only be attributed to chance since it is logically impossible for a 10-trial criterion to be more difficult than a 20-trial criterion.

responses) made *more* perseverative errors than did the groups with fewer criterional trials. The 20-trial criterion group made fewer perseverative errors than the 10-trial criterion group. Little can be said about these results since there was a significant Degree of Learning × Percentage Reinforcement interaction and it is there that we must look for any answers. The same can be said for percentage reinforcement. Although 25% reinforcement resulted in a greater number of perseverative errors, the interaction is the important determinant of the results.

Under 100% reinforcement, the results were clearly the same as those of other studies cited in the introduction: in general, the fewer the criterional trials, the greater the number of perseverative errors. In the study by Grant and Cost (1954) it was also shown that there was little difference between the 20-trial criterion group and the 40-trial group. This same result was shown here.

However, under 50% reinforcement the opposite results were obtained. Although the effect is slight, the greatest number of perseverative errors was made by the groups required to reach a 40-trial criterion, the second most frequent number of perseverative errors was made by the 20-trial group, and the least number of perseverative errors was made by the group required to reach a 10-trial criterion. This result could be explained in terms of the hypothesis stated in the introduction. That is, if the reinforcement pattern in Stage I is similar to that in Stage II, the change would be difficult to detect and more perseverative errors would be made. Furthermore, under partial reinforcement, a greater number of trials in Stage I would increase the length of runs of nonreinforced trials which would be learned by *S* not to be infirmatory in nature. Thus the pattern of reinforcement for the same response in the second stage (0%) would be more similar to the first stage pattern for those groups with greater numbers of criterional trials. This, then, results in slower extinction, or a greater number of perseverative errors.

Somewhat similar results as in the 50% groups were obtained for 25% reinforcement. After 40 criterional trials there was clearly a larger number of perseverative errors than after the other criteria. This fits into our analysis. However, the 10-trial criterion group made more perseverative errors than the 20-trial criterion group. No explanation could be found for this latter result.

Although predictions were made for the effect of various criteria within the three percentage reinforcement groups, nothing was said of the relative position of all nine groups with respect to each other. If it were supposed that the effect of the percentage of reinforcement on extinction, as it is normally found, were so strong that it applied under all conditions of degree of learning, it might have been predicted that the groups—no matter what the degree of learning was—would rank 100%, 50%, 25% going from fewest to most perseverative errors. This was clearly not the case. In addition to this possible prediction, it might have been expected that the difference between the various percentages of reinforcement would increase, the greater the number of criterional trials. This again was not found.

Two methodological factors in the present experiment may account for the unexpected results. One of these factors was that all *S*s received the same instructions. Thus, even the 100% groups received partial reinforcement instructions. These instructions were given so that the only differences found in the experiment would be due to the experimental conditions alone and not due to differences in instructions. For 100% reinforcement, though, these instructions may have caused *S*s to make more perseverative errors than they would normally tend to make. Unfortunately the values in this experiment cannot be directly compared to those found by Grant and Cost (1954) since they used the full WCST deck of cards and only 24 of these were used in this experiment.

The other methodological factor was

the impossibility of determining in advance when the last criterional response would be made by any *S* in the partial reinforcement groups. Therefore it was impossible to reinforce the last criterional trial. Since this was the case, there was no definite demarcation between Stage I and Stage II in the partial reinforcement groups. For these groups, especially for the 25% groups, extinction therefore actually started before the last criterional trial was reached. Thus the curves for the 50% groups and especially for the 25% groups are depressed, and more perseverative errors were made than is shown by the results.

It was shown that *S*s who sorted according to color in Stage II made more nonperseverative errors than *S*s who sorted according to form in Stage II, when both groups sorted according to number in Stage I. This result is in agreement with the results of Grant and Curran (1952) which show that at Stage II more nonperseverative errors are made when sorting under color than when sorting under form.

Summary

An experiment relating perseveration to percentage of reinforcement and degree of learning was performed using a modified form of the Wisconsin Card Sorting Test (WCST). The results showed that there was a significant interaction between these two variables when perseverative errors were considered. Under 100% reinforcement, the greater the degree of learning the less the perseveration (within limits). Under partial reinforcement, somewhat of an opposite result was found indicating that the greater the degree of learning, the greater the number of perseverative errors. These results were discussed in terms of the similarity of the pattern of reinforcement during the acquisition stage and the nonreinforcement of the same response during the extinction stage. Supposedly, similarity leads to generalization and thus, slower extinction (greater perseveration).

The learning data showed that the more stringent the criterion, the more difficult the solution of the problem was, though this held true only for the most stringent criterion. Percentage reinforcement also affected the learning time in that the greater the percentage of reinforcement the simpler was the task.

References

Capaldi, E. J., & Stevenson, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

Erlebacher, A. Perseveration as a function of degree of learning and percentage reinforcement in card sorting. Unpublished master's thesis, University of Wisconsin, 1960.

Grant, D. A., & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.

Grant, D. A., & Cost, J. R. Continuities and discontinuities in conceptual behavior in a card sorting problem. *J. gen. Psychol.*, 1954, **50**, 237–244.

Grant, D. A., & Curran, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

Humphreys, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

Pubols, B. H. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

(Received August 25, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 518–522

SHORT-TERM RETENTION AS A FUNCTION OF AVERAGE STORAGE LOAD AND AVERAGE LOAD REDUCTION [1]

LYNE STARLING REID, KENNETH E. LLOYD,[2] H. RAY BRACKETT, AND WILLIAM F. HAWKINS [3]

University of Virginia

A series of experiments (Lloyd, Reid, & Feallock, 1960) has been previously reported in which short-term retention was studied in a task that required *S*s to recall sequentially presented English words. The task differed from the one used to study classic memory span in that the presentation of items and their recall were not discretely separated. Instead, both occurred in a variety of serial orders over an extended period of time. Each item to be remembered was followed at some point in the sequence by a cue which specified that *S* must recall the item. The recall point may have followed immediately the presentation of the item, or may have been separated from it by several other items, by other recall points, or by both. Throughout the task *S* was presented with new words he had to remember and from time to time was required to recall words previously presented. Recall errors in the task were found to be related in a systematic manner to the average number of items *S* was required to remember. Furthermore, if the items are familiar, *S* appears to be equally capable of storing items from different sized populations so long as the average number of items to be remembered, i.e., the average storage load, remains constant.

In these earlier studies *S* recalled only one item at each recall point. In the studies to be reported the generality of average storage load is assessed for sequences in which *S* is required to recall more than one stored item at a recall point. As in the original experiments the words to be remembered are instances of a word class and the class names serve as recall points. The following series of words is an example of such a sequence: Berlin, Willow, Moscow, Dartmouth, Elm, CITY, Golf, Ford, Soccer, TREE, COLLEGE, ... When the word CITY occurs *S* should respond with "Berlin" and "Moscow." Similarly, at TREE the correct recalls are "willow" and "elm." At COLLEGE only one item, "Dartmouth," is appropriate.

The above sequence can be used to make clear the concept of average storage load and a new summarizing measure *average load reduction*. At the recall point CITY, *S* has been presented five items, and, thus, has a storage load of 5. Upon the recall of "Berlin" and "Moscow" this is reduced by 2, but before TREE occurs *S* is presented three further items increasing the load to 6. Again at TREE the load reduction is 2, so to this point average storage load is 5.5 and average load reduction 2.0. In the present experiments these averages were computed over much longer runs, but were obtained in like manner by simply averaging the load and reduction values at each of the recall points in a sequence.

Two experiments are reported. The first determines the effect of various values of average storage load upon the recall of sequentially presented items when average load reduction is held constant, but at a value greater than 1. In Exp. II average load reduction is varied while average storage load is held constant.

[1] The reported studies were supported by Contract No. DA-49-007-MD-537 between the University of Virginia and the Office of The Surgeon General, Department of the Army for which Richard H. Henneman is principal investigator.

[2] Now at Washington State University.

[3] Now at George Peabody College.

EXPERIMENT I

Method.—The design of the experiment parallels that of Exp. I in Lloyd et al. (1960). The *Ss* stored and recalled words presented to them aurally by a magnetic tape recorder. In an experimental session they listened to two sequences. Each sequence was composed of 12 subsequences constructed to have any one of four values of average storage load: 2.5, 3.5, 4.5, and 5.5. Within a sequence each value of average storage load appeared three times in a partially counterbalanced order.

Average load reduction in all sequences was 2. This was accomplished by using individual reductions of either one, two, or three stored items at each recall point. A reduction of 2 was the most frequent value.

Throughout a sequence the values of the storage loads and load reductions at individual recall points occurred in a variety of combinations with the exception of the 2.5 subsequences. When average load and average reduction are nearly equal as is the case in the 2.5 subsequences, the individual loads and reductions must also be nearly equal. For this reason, in the 2.5 subsequences the load and reduction were equal at one-half the recall points. At these points *S* had to recall all the items being stored. Since it is impossible to avoid a reduction of load to 0 in the 2.5 subsequences if an average load reduction of 2 is to be maintained, reductions to 0 were introduced into the subsequences of higher storage loads even though they are unnecessary.

All subsequences contained 20 items which *S* had to remember and for which he was scored. With an average load reduction of 2, this means there were 10 recall points for these items. It was necessary, however, to insert extra items in order to maintain the desired loads at recall points near the ends of the subsequences. These extra items were called for immediately after the tenth recall point in the subsequence and were not scored.

The items were instances of classes of English words, and the class names were used as recall points. The eight class names were AUTOMOBILE, TREE, CITY, METAL, SPORT, CLOTHING, COLLEGE, and BIRD. Each class contained 8 items judged to be familiar to *Ss*. However, since the words were probably not equally familiar, the *Ss* were given preliminary familiarization with the recall material at the beginning of the first experimental session. They were presented with a list of the 64 items grouped into the eight classes and were allowed 5 min. to become familiar with the items. In addition, they were permitted to keep the list for reference during the first experimental session and prior to each subsequent session were required to study the list for 3 min.

Four sequences were recorded on the magnetic tape. The sequences differed in the serial order of the 12 subsequences, and in the particular words that were substituted at the different loads. A presentation rate of one item or recall point every 3 sec. was used.

The 12 *Ss*, University of Virginia undergraduates, were instructed to remember each item as they heard it on the tape and to recall the items when the appropriate class name was heard. It was pointed out that one, two, or three words might be required at any recall point, and that any substitution, omission, or addition would be scored as an error. They were encouraged to guess if uncertain about a recall. Each *S* responded to two sequences a day for 4 days.

TABLE 1

MEAN ERRORS AS A FUNCTION OF AVERAGE STORAGE LOAD DURING THE 4 DAYS OF PRACTICE

Day	Average Storage Load							
	2.5		3.5		4.5		5.5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	1.94	0.95	4.05	1.40	6.76	1.25	8.33	1.59
2	0.57	0.41	1.65	1.21	5.22	1.98	6.47	1.57
3	0.44	0.41	0.83	0.74	3.97	1.31	5.04	1.34
4	0.10	0.05	0.47	0.44	3.01	1.16	4.93	1.29

Results.—All *S*s on each of the 4 days responded to six subsequences at each value of average storage load. The means of errors over these repeated measures are presented in Table 1. As in the previous studies (Lloyd et al., 1960) recall errors increase systematically with increases in average storage load. For the analysis of variance over the 4 days of practice a square-root transformation of the data was made since the errors for low storage loads are skewed. From this analysis storage load, practice, and the Load × Practice interaction are found to be significant beyond the 1% level.

Since the recall scores on Day 4 are considered the best estimate of stable performance a second analysis was performed. Table 2 summarizes the trend analysis of transformed mean error scores for each *S* on this day. The overall trend is highly significant. Although the first three components are significant, 95% of the overall-trend sum of squares can be attributed to the linear component. The significant quadratic and cubic components probably arise from the relatively small error differences at the lower storage loads compared to the much larger error difference beyond the 3.5 load.

TABLE 2

TREND ANALYSIS OF TRANSFORMED SCORES OF AVERAGE STORAGE LOAD

Source	*df*	*MS*	*F*
Overall trend	3	6.42	214.00*
Linear	1	18.29	457.25*
Quadratic	1	0.22	11.00*
Cubic	1	0.75	25.00*
Between individual means	11	0.16	5.33*
Between individual trends	33	0.03	
Linear	11	0.04	
Quadratic	11	0.02	
Cubic	11	0.03	
Total	47		

* $P < .01$.

EXPERIMENT II

Method.—Values of average load reduction are obviously dependent upon values of average storage load; there can never be greater reduction than there is load. It was for this reason in Exp. I where storage load was varied that load reduction was held constant. Storage load was in Exp. II held constant at 4.5, while average load reduction was varied through six values: 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5.

Two subsequences were constructed for each value of load reduction. The 12 subsequences were similar to those of Exp. I; their length, however, was increased to 35 items in order to provide adequate sampling of high-reduction values. Furthermore, the number of recall points varied from subsequence to subsequence. This was necessarily the case since the number of recall points in a subsequence decreases as average load reduction is increased. There were 35 recall points in the 1.0 subsequences, 23 at 1.5, 18 at 2.0, 14 at 2.5, 12 at 3.0, and 10 at 3.5.

Four sequences were prepared using the items and class names of Exp. I. The sequences differed in the serial order of the six subsequences and in the words assigned to the subsequences. Presentation rate for the items was one every 3 sec., and the duration of each recall point was 4 sec.

Instructions to *S*s were similar to those of Exp. I. Ten University of Virginia undergraduates responded to two sequences a day during 4 days of practice.

Results.—On each of the 4 days all *S*s responded to two subsequences at each value of average load reduction. The mean errors on these repeated measures were averaged for each *S* at each value of average reduction and are shown in Table 3. It is apparent that when average storage load is held constant recall errors systematically decline as average load reduction is increased. An analysis of variance over the 4 days of practice reveals load reduction

TABLE 3

MEAN ERRORS FOR AVERAGE LOAD REDUCTION DURING THE 4 DAYS OF PRACTICE

Day	Average Load Reduction											
	1.0		1.5		2.0		2.5		3.0		3.5	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	21.70	2.13	14.65	2.65	12.75	3.56	9.30	3.17	7.75	3.61	5.65	3.13
2	17.55	3.46	13.50	3.39	8.90	2.95	7.15	3.01	5.35	3.32	4.00	3.42
3	15.85	2.73	10.30	3.83	6.40	2.46	4.95	2.52	3.95	2.67	2.30	2.00
4	14.75	3.54	9.90	3.56	5.90	3.44	4.85	2.71	3.65	2.99	2.30	2.13

and practice to be significant beyond the 1% level; the Reduction × Practice interaction was not significant.

Table 4 contains a trend analysis of the data on Day 4. Since the errors for the higher values of load reduction are skewed a square-root transformation of the data was made. The overall trend, and both the linear and quadratic components are significant beyond the 1% level. Again, as in Exp. I, most of the variance in the overall trend is accounted for by the linear component.

TABLE 4

TREND ANALYSIS OF TRANSFORMED SCORES OF AVERAGE LOAD REDUCTION

Source	*df*	*MS*	*F*
Overall trend	5	8.20	82.00*
Linear	1	39.21	261.40*
Quadratic	1	1.43	20.43*
Cubic	1	0.17	2.83
Quartic	1	0.08	—
Quintic	1	0.13	—
Between individual means	9	2.50	25.00*
Between individual trends	45	0.10	
Linear	9	0.15	
Quadratic	9	0.07	
Cubic	9	0.06	
Quartic	9	0.10	
Quintic	9	0.14	
Total	59		

* $P < .01$.

DISCUSSION

In these studies *S* stored and recalled words in a single sequence for a period of 15 to 20 min. During this time the words and their recall points occur in a variety of combinations which introduce many variable conditions that can affect performance. For example, the many different patterns of presentation and recall produce variations which have, by analogy to long term retention paradigms, proactive and retroactive effects. In this regard, however, there are two important differences between the way these conditions vary in the present studies and the manner they typically vary in a long term retention study. First, in the long term retention studies a relatively few values of one or a small number of variables are manipulated in a systematic fashion while the attempt is made to hold all other conditions constant. In the two studies under consideration, on the other hand, many values of a number of variables are permitted to vary in a nonsystematic fashion. Second, in long term retention studies conditions usually can be varied in an independent manner; within the sequences used in the present studies, however, prior and interpolated items and recall points cannot be manipulated independently of one another. A given pattern of items and recall points restricts to a considerable degree the previous and later occurrence of other items and recall points.

Both these factors make highly desir

able variables which summarize the variety of conditions and dependencies within the task. The earlier experiments of Lloyd et al. (1960) established the fact that, in tasks requiring the recall of only one item at each recall point, the average number of items *Ss* are required to remember is related in a reliable manner to their recall errors. Experiment I extends the generality of these findings to situations in which the recall of more than one item at a recall point may be required. Furthermore, the highly significant and consistent differences found between the different values of average storage load point up the sensitivity of performance to this variable. As was pointed out above, *Ss* are storing and recalling words in a sequence for a period of 15 to 20 min. During this period load changes occurred approximately every 1.5 min., and recall paralleled these changes throughout the sequence.

Experiment II demonstrates that a second important summarizing variable is significantly related to performance in such tasks. At first it may seem surprising that improved retention is related to increases in average load reduction, but a consideration of the variables contributing to its effect makes such an outcome reasonable. Two important changes took place in Exp. II with increases in load reduction. The first, a decrease in number of recall points, is simply a correlated condition of the sequential task. This decrease in recall points does, however, reduce the average amount of interpolated activity between the presentation of items and their recall, and may account, in part, for the improved performance. The second, the fact that increased load reduction means more items from a single class are being stored, provided the opportunity for the grouping of related items. In this latter respect average load reduction is obviously related to the phenomenon of clustering (Bousfield, 1953) and to Miller's (1956) concept of the chunk.

Although the reduction in recall points and the opportunity for grouping are in a sense confounded in the present study, they can be varied independently. If, for example, arbitrary labels are used to designate the categories into which words are grouped for recall, the extent to which the words grouped together are related can be varied over a wide range of values. The pattern of their presentation and recall, on the other hand, can be held constant by merely substituting the different words into the same sequences. An investigation of the influence of relationship among grouped items upon performance is under way. Such an investigation points up one of the major advantages of the present method, i.e., a number of complex conditions can be held constant while one or more experimental variables are systematically manipulated.

Summary

Two experiments are reported in which short-term retention was studied in a situation requiring *S* to remember familiar English words until their recall was requested. During the time *S* was required to remember these items he was presented other words or asked to recall previously presented items. Experiment I demonstrated that in a task where the recall of more than one item could be requested at a given point in the sequence, errors increased systematically as the average number of words to be remembered over blocks of recall points, average storage load, was increased. In Exp. II it was found that when the average number of items to be remembered was held constant, errors decreased as the average number of items to be recalled at given recall points, termed average load reduction, was increased.

References

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

Lloyd, K. E., Reid, L. S., & Feallock, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

(Received August 26, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 523–531

THE EFFECTS OF INDUCED MUSCLE TENSION DURING TRACKING ON LEVEL OF ACTIVATION AND ON PERFORMANCE [1]

LAWRENCE R. PINNEO [2]

Allan Memorial Institute, McGill University

Since Courts' (1942) review, several experiments have been reported on the effects of induced muscle tension on performance, but to the author's knowledge none have been concerned with the basis of the relationship. In a theoretical paper Meyer (1953) suggested that the effects of induced muscle tension on performance are the result of simultaneous responses occurring at the level of the motor cortex. Though a useful concept, Meyer fails to consider the involvement of the reticular activating system and level of arousal in muscle tension phenomena (Delafresnaye, Adrian, Bremer, & Jasper, 1954).

In contrast to Meyer (1953), Malmo (1959, p. 370) suggests that induced tension may be one of the ways in which activation level can be varied, and that activation level, in turn, is a function of the reticular activating system. If induced tension is indeed a reliable means of varying activation level, then in addition to behavioral effects, induced tension should also produce regular and consistent changes in the various physiological indicants of activation. Experiments by Freeman and Simpson (1938) encourage this idea. They have shown that palmar conductance increases as a function of muscle tension induced in the legs, and Freeman (1938) has shown that palmar resistance changes linearly with induced muscle tension during performance. However, these experiments were unfortunately limited to the one physiological measure of palmar conductance, while Malmo (1959) and Schnore (1959) have shown the need for a wide coverage of physiological functions in studies of activation. The purpose of this experiment was the exploration of the relation between induced muscle tension, activation, and level of performance.

METHOD

Subjects

The *Ss* were 38 McGill University undergraduate and graduate males, ranging in age from 18 to 30 yr. All were paid $5.00 for the 3-hr. session. Only right-handed *Ss* with normal hearing were used.

[1] This paper is part of a thesis presented to the Faculty of Graduate Studies of McGill University for a PhD degree. It was supported in part by the Medical Research and Development Division, Office of The Surgeon General, Department of the United States Army: Contract Number DA-49-007-MD-626; Defence Research Board, Department of National Defence, Canada: Grant Number 9425-04; National Institute of Mental Health, National Institutes of Health, United States Public Health Service: Grant Number M-1475; National Research Council of Canada: Grant Number A.P. 29; and Department of National Health and Welfare (Canada): Grant Number 604-5-69.

The author expresses his thanks to Robert B. Malmo for his continuing interest, guidance, and support; D. J. Ehrlich, P. F. MacNeilage, and L. D. Rust for their discussion and criticism of the experiment; W. Mundl and J. F. Davis for assistance with the apparatus; and Jeanne M. Pinneo for her assistance in carrying out the experiment.

[2] Now at the Institute of Physiology, University of Pisa, Italy.

Apparatus[3]

Muscle tension induction system.—A modified hand dynamometer was used for inducing tension. When the spring steel of the dynamometer was compressed, two strain gauges mounted on each side of the spring were deformed. This resulted in a change of the electrical resistance of the wire in the strain gauge proportional to the strain or the grip on the dynamometer. Each strain gauge was an arm of a balanced Wheatstone bridge. Changes in the resistance of either or both of the strain gauges upset the electrical balance of the bridge. The change in signal due to this imbalance was fed into a push-pull, strain gauge amplifier.

Part of the amplified signal was used to continuously monitor and permanently record *S*'s instantaneous performance, and part was fed back to *S* to enable him to maintain his level of pull within prescribed limits. These limits were set by an ac microammeter relay equipped with two movable contacts. Both contacts could be positioned anywhere on the meter face and their relative positions could also be varied. By calibrating the meter deflection in kilograms, and using each of the meter-relay contacts as a switch to an audio oscillator, "low" and "high" tension limits could be determined. When *S* did not squeeze the dynamometer sufficiently hard, he heard a 600-cycle tone via one contact, while if he squeezed too hard he heard an 800-cycle tone via the other contact. The desired pressure was obtained when *S* gripped the dynamometer such that he heard no tone at all.

Auditory tracking system.—This system was similar to that described by Davis, Stennett, and Quilter (1957), except that tracking was done with the right foot instead of with the hand. In addition the function generator used in this system consisted of a potentiometer bridge circuit whose imbalance signified error in direction and intensity. The *S*, who was blindfolded for purposes of EEG recording, sat in a semireclining chair and performed a simple tracking task in the following manner. He first depressed the foot pedal until it came against a mechanical stop after a movement of 8° of arc. He next allowed the pedal to return to its normal position at the same rate at which the pedal was depressed. There was a stop at the upper, as well as at the lower terminal position of the pedal, and in order to entirely avoid error it was necessary for the foot to move off each stop immediately. In tracking, a constant rate of movement was required. If the rate of pedal movement down and up was in exact conformance with that of the function generator, *S* was "on target" and he heard no sound in his earphones. Failure to keep pace with the function generator was signaled to *S* by means of a 1000-cycle tone in his earphones. The loudness of this tone varied as a function of *S*'s "distance off target," and direction of error was cued by whether the tone appeared in the right or the left earphone. "Down" errors (i.e., pressing down on the pedal at too fast a rate or lagging behind on the way up) were indicated by the tone appearing in the earphone on the right; conversely, "up" errors were indicated by the tone appearing in the earphone on the left.

The number of up and down errors made per trial and the total distance *S* was off target (DOT) were the measures of performance. Number of errors were recorded on two electrical impulse counters, and DOT was recorded graphically by means of an electronic integrator with a 4-sec. discharge rate (Davis, 1959). This latter system was calibrated such that a certain distance off target in degrees produced a given deflection on the ink writer, which was then measured with special scales.

Physiological recording apparatus.—Three chart drives with ink-writing galvanometers were used for recording physiological tracings and error data. The primary record for electromyograms (EMG) from active and passive muscles, two EEG measures, and the heart rate (EKG), were recorded on a standard Grass IVA electroencephalograph (paper speed 25 mm/sec). Signals taken from the output of the driver stages of the Grass were led into electronic integrators which summated the muscle potentials over successive 4-sec. periods, and were recorded as deflections on a second chart drive. The EEG was quantified by passing these signals through band-pass filters (Ross & Davis, 1958) with cutoffs to provide bands of 8–12 cycles/sec and 18–27 cycles/sec. Outputs of the filters were also integrated and recorded on Chart Drive 2 (paper speed: 1 mm/sec).

[3] Much of the apparatus used in this experiment was designed and built in this laboratory. Detailed specifications and circuit diagrams of the tracking apparatus, the strain gauge tension induction system, the EEG band-pass filters, and the electronic integrators are on file at the Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, McGill University, Montreal, Canada.

Chart Drive 3 (paper speed: 2.5 mm/sec) was used to record respiration (obtained with a Phipps & Bird pneumograph and recording tambour), palmar conductance, DOT, and the output of the strain gauge amplifier.

Recording electrodes.—All electrodes except those used for palmar conductance were made of cellulose sponge of 1 cubic in. Heart rate was recorded from electrodes placed on the chest wall immediately below the heart, and on the right shoulder. Electrodes for EMG recording were placed on the extensor muscles of the right and left forearms and the calves of the right and left legs (Davis, 1959). Two bipolar EEG leads were taken from the nondominant hemisphere from frontal and occipital positions equivalent to placements C2-F2 and P2-02 of the 10/20 system used at the Montreal Neurological Institute (Jasper, 1941), frontal leads being connected with the EEG lead to the 18–27 cps filters, and occipital leads connected with the 8–12 cps filters.

Palmar conductance (PC) was recorded from the left palm by means of a monopolar method described by Malmo and Davis (1961).

Other apparatus.—The *S*'s chair and tracking apparatus were located in a separate room from that of the amplifying and recording devices. It was a relatively soundproof room, entirely shielded with copper screening to reduce electrical interference, and was supplied with an air conditioning plant for the maintenance of constant temperature. Separating *S*'s room from *E*'s control room was a double wall with a screened observation window. A two-way intercommunication system was employed for instructing *S*s and answering questions.

Procedure

All instructions pertaining directly to the experiment were read to each *S*. When *S* first arrived he was informed that the experiment was designed to study the effects of induced muscular tension on his ability to perform a simple psychomotor task, and that the effects of this tension would be determined by simultaneously measuring certain physiological variables and the muscle tension directly. An assistant then prepared *S* by placing the electrodes as discussed above.

The *S* was then instructed by *E* in the method of inducing tension and in the method of maintaining a given level of tension. Subjective levels of tension were then determined by having *S* squeeze the dynamometer very lightly for 30 sec. After a rest period, this procedure was repeated twice more. The average reading of his pull in kilograms at the end of the three, 30-sec. periods was designated *S*'s Very Light (VL) tension. After another rest period, *S* was asked to squeeze the dynamometer as tightly as he could and to hold this for 30 sec. The reading at the end of the 30 sec. was designated his Maximum (Max.) pull. Following another rest period, *S* was required to subjectively bisect the apparent pull on the dynamometer of the previous two instances (Very Light and Maximum), and to hold this for 30 sec. This also was repeated twice more, and the average of the three dynamometer pulls in kilograms at the end of the 30-sec. periods was designated as *S*'s Very Heavy (VH) tension. Similarly, a Medium (M) tension was determined by bisecting the apparent pull between the Very Light and the Very Heavy, as determined by the average of three readings. Light (L) and Heavy (H) tension conditions of pull were then arbitrarily chosen as the physical midpoints, in kilograms, between the Very Light and Medium, and the Medium and Very Heavy, respectively.

A headset with independent earphones for tracking was placed on *S*'s head and adjusted for comfort. The *S* was then instructed by *E* in the tracking procedure as described above. Each trial consisted of two complete cycles of the tracking pedal being depressed and released. This required 96 sec.; 96 sec. rest was also given between each trial. One practice trial was followed by the learning trials to criterion. The criterion was the tracking of 3 successive trials without improvement, but never less than 6 nor more than 10.

A 10-min. rest period and a glass of water were given between learning and tension trials. No smoking was allowed. During the rest period all equipment was again checked to conform with calibration standards.

The five tension conditions of (VL, L, M, H, and VH) were given once only to each *S* in an order determined by a table of random numbers. The only difference between the tension trials and the learning trials was that during the entire trial *S* was required to maintain the particular tension on the dynamometer in addition to tracking. After two practice trials, *S*s had no difficulty in simultaneously tracking and maintaining a steady level of dynamometer pull.

Two control conditions of No Tension and Exertion (EX) were also randomly presented with the tension conditions. The No Tension control did not differ from the learning trials.

Its purpose was to gauge the effects of performance with no tension. The Exertion condition was introduced in an attempt to determine the effect of induced tension alone (without tracking) on the various physiological measures. This condition required each of 8 *Ss* to squeeze the dynamometer at a pressure equivalent to his Very Heavy pull, and each of 25 *Ss* to squeeze the dynamometer at a pressure equivalent to his Maximum pull. Five *Ss* did not have an Exertion condition at all. Besides squeezing the dynamometer, in the Exertion condition each *S* was required to push the foot pedal down and keep it there for the length of the trial. He also heard a tone in both ears equal in intensity to what he would have heard had he been off target two degrees. Physical exertion was thus at a high level. The primary difference between this and other trials was that there was no actual tracking performance.

To increase the rate of learning during the early part of the experiment, *Ss* were told their score in terms of time off target at the end of each trial. This was not continued during the tension trials, however, in order to prevent associations of a particular score with a given level of tension and consequent excitement or activation if the score were relatively poor.

The mean tolerance allowed for the maintenance of tension was about 1.5 kg. In the pilot experiments and the main study, it was found that after a little practice this tolerance was more than sufficient to enable *Ss* to maintain tension at quite steady levels over an entire trial.

Treatment of Data

EEG.—Outputs from the band-pass filters were integrated with 4-sec. discharge rates and recorded as deflections (Malmo & Surwillo, 1960, p. 7). During the 96-sec. trial, 24 of these deflections were recorded, and the level of EEG activity for a given trial was then defined as the mean voltage of the 24 integrated deflections.

EKG.—EKG was obtained by counting the beats in each successive 12-sec. period and multiplying by 5 to get the mean number of beats per minute for that 12-sec. period. The eight values thus calculated were averaged for the mean number of heart beats per minute for the trial.

EMG.—EMG being less variable than EEG it was satisfactory to sample only eight peaks during a trial. The measured peaks for the right leg coincided with the deflection occurring approximately midway up and down the rises and falls in muscle tension due to tracking (Deflections 3, 4, 9, 10, 15, 16, 21, and 22 were measured). Deflections were measured with special scales calibrated in microvolts, and the level of muscle activity for a trial was then defined as the mean voltage of the eight measured deflections.

Respiration and palmar conductance.—Respiration rate was found by counting the number of complete inspiration-expiration cycles from one 12-sec. line to the next, correct to the nearest ¼ cycle. The mean rate for the trial was obtained by multiplying this value by 5 to get the number of cycles per minute for that period, and then obtaining the mean of the eight values. Palmar resistance was measured at 12-sec. intervals on the tracing with a special scale calibrated in ohms. This resistance value was converted to its reciprocal, conductance, and the average conductance value in micromhos for the trial was obtained by calculating the mean of the eight conductance measures.

DOT and strain gauge.—The integrated deflections of Distance Off Target and the pull on the strain gauge dynamometer (SG) were measured in the same way as the EEG. That is, each of the 24 deflections was measured using a special plastic scale, calibrated in degrees for DOT and kilograms for SG. The average of the 24 deflections was taken as the mean DOT or grip pressure for the trial.

Errors.—The number of "up" and "down" errors per trial was obtained directly by means of two electromechanical counters.

Results

Figure 1A shows the continuous and approximately linear rise in force, as measured by the strain gauge dynamometer, through the graded series of induced tensions. Parts B and C of Fig. 1 show that the effect of induced tension in this experiment was to impair tracking performance, with the higher tensions having a more detrimental effect than the lower ones. A two-way analysis of variance of these data shows that the between-trials variance of errors as measured by Distance Off Target was significant ($P = .05$) for the induced tension conditions.

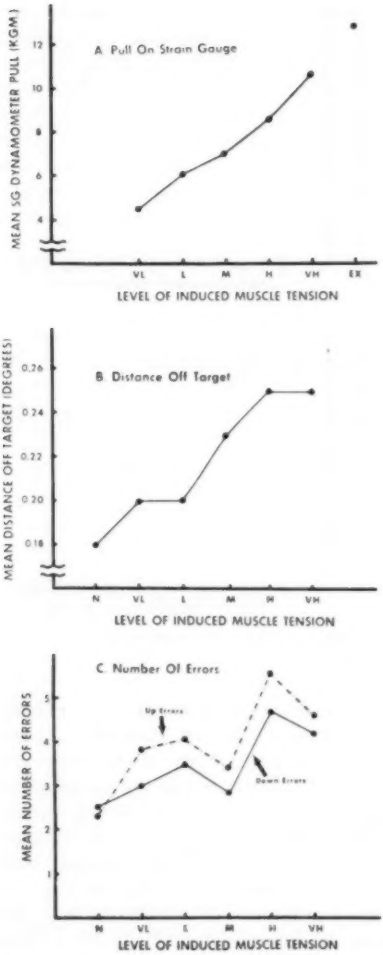

FIG. 1. Performance measures as a function of induced muscle tension.

Figure 2 presents the EMG data. As in the case of the strain gauge, the curve for muscle tension in the right arm, as a function of the amount of tension induced, is approximately linear. The other three curves, for muscles not directly involved in pulling on the dynamometer, likewise show incremental changes corresponding to increasing values of induced tension. In Fig. 2B the right leg is designated as active because it was used in tracking, but the load on the leg muscle was constant, of course, from one induced tension condition to the next. The two left limbs (Fig. 2C and 2D) were not engaged in any activity. Despite this passivity, however, their curves show steady rise through the ascending series of induced tension values up to Cond. VH (i.e., through the series of tensions that were induced during tracking). Analysis of variance for all four muscles also showed the between-trials variance to be significant ($P < .01$) over the range of tension conditions with tracking.

The drop in the left arm in tension from the Cond. VH during tracking to Cond. EX (no tracking) should be especially noted in Fig. 2C. This is shown as a point on the far right. The drop was significant ($P < .01$). It will be recalled that 33 of the 38 *S*s were included in Cond. EX. Twenty-five of these *S*s pulled on the dynamometer their maximum amount during this trial, while 8 *S*s gripped with the same pressure as they used in Cond. VH. With this amount of exertion, the right leg EMG (Fig. 2B) showed a significant rise from VH to EX (the right leg held the tracking pedal all the way down during this condition, exerting the maximum leg tension for a trial). Yet, the right arm EMG (Fig. 2A) which was squeezing the dynamometer, shows only a slight and very nonsignificant rise. In fact, 48% of the 33 *S*s actually showed a fall in the right arm EMG from the VH tracking condition to the EX nontracking condition; 75% of the 33 *S*s showed a fall in the left arm EMG from VH to EX. The slight rise shown in Fig. 2D is not significant,

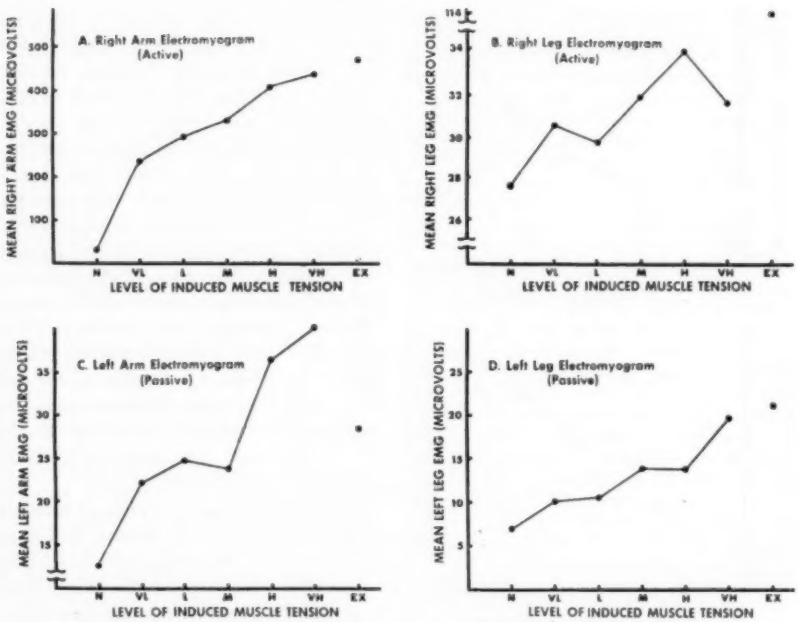
A. Right Arm Electromyogram (Active)
MEAN RIGHT ARM EMG (MICROVOLTS)
LEVEL OF INDUCED MUSCLE TENSION
B. Right Leg Electromyogram (Active)
MEAN RIGHT LEG EMG (MICROVOLTS)
LEVEL OF INDUCED MUSCLE TENSION
C. Left Arm Electromyogram (Passive)
MEAN LEFT ARM EMG (MICROVOLTS)
LEVEL OF INDUCED MUSCLE TENSION
D. Left Leg Electromyogram (Passive)
MEAN LEFT LEG EMG (MICROVOLTS)
LEVEL OF INDUCED MUSCLE TENSION
N VL L M H VH EX

FIG. 2. Muscle tension of active and passive limbs as a function of induced muscle tension.

of course. A summary of the percentage of *S*s whose measures of activation fell for Cond. EX as compared with Cond. VH is shown in Table 1.

Similar curves for palmar conductance, heart rate, and respiration are presented in Fig. 3. It is noteworthy that there are no reversals in the curves which all show a regular progressive rise through the increasing series of induced tensions. Again, a two-way analysis of variance revealed the between-trials variance of all three measures to be significant ($P < .01$) over the range of induced tension. The fall in palmar conductance from VH to EX in Fig. 3A was significant ($P < .05$) as was the rise in heart rate ($P < .01$). The

TABLE 1

PERCENTAGE OF SUBJECTS WITH LOWER VALUES FOR PHYSIOLOGICAL RECORDINGS UNDER THE EXERTION CONDITION THAN UNDER THE TRACKING COND. VH

S's Pull on EX	*N*	Right Arm	Right Leg	Left Arm	Left Leg	Palmar Cond.	Heart Rate	Resp.	EEG 8–12	EEG 18–27
Max.	25	48.00	32.00	72.00	40.00	16.00	0.00	56.00	52.00	60.00
VH[a]	8	50.00	12.50	87.50	62.50	62.50	62.50	75.00	62.50	12.50
Comb.	33	48.48	27.27	75.75	45.45	27.27	15.15	60.60	54.54	48.48

[a] This VH condition is not to be confused with the VH condition during tracking. These 8 *S*s took their Exertion (i.e., no tracking) condition by pulling the same kilogram values on the dynamometer as they had pulled in the VH tracking condition.

slight change in respiration was not significant.

Finally, Fig. 4 presents the EEG data which again show remarkable correspondence with the increasing levels of induced tension. Again, both measures showed significant between-trials variance over the range of tension conditions by a two-way analysis of variance (8–12 cps, $P = .05$; 18–27 cps, $P = .01$). The large reversal between Cond. VH and EX is especially to be noted (i.e., falling 8–12 cps amplitude, significant with $P < .01$, compared with rising 18–27 cps amplitude almost significant at the .05 level).

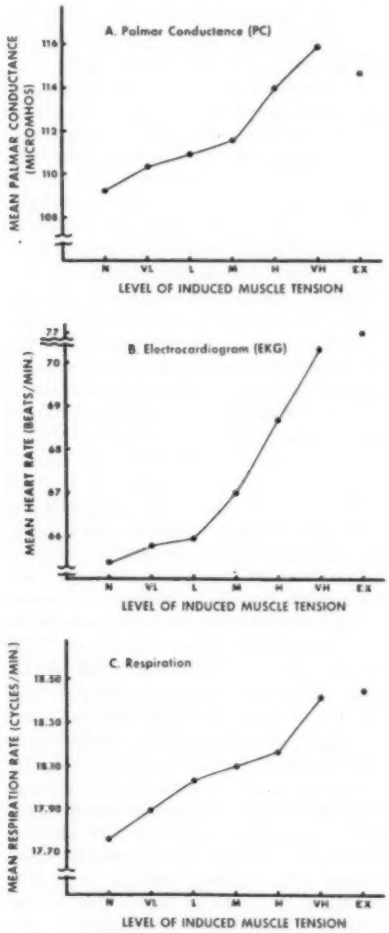
A. Palmar Conductance (PC)
MEAN PALMAR CONDUCTANCE (MICROMHOS)
LEVEL OF INDUCED MUSCLE TENSION
B. Electrocardiogram (EKG)
MEAN HEART RATE (BEATS/MIN.)
LEVEL OF INDUCED MUSCLE TENSION
C. Respiration
MEAN RESPIRATION RATE (CYCLES/MIN.)
LEVEL OF INDUCED MUSCLE TENSION

FIG. 3. Peripheral measures of autonomic activity as a function of induced muscle tension.

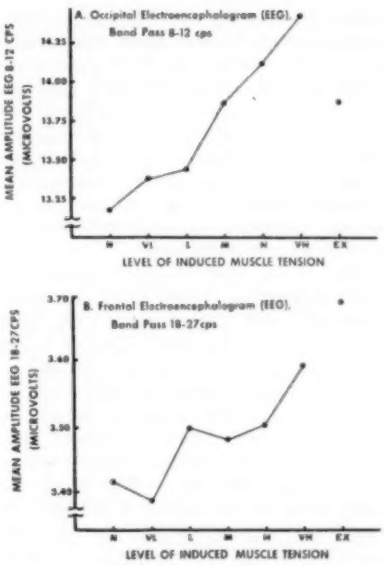
A. Occipital Electroencephalogram (EEG).
Band Pass 8-12 cps
MEAN AMPLITUDE EEG 8-12 CPS (MICROVOLTS)
LEVEL OF INDUCED MUSCLE TENSION
B. Frontal Electroencephalogram (EEG).
Band Pass 18-27cps
MEAN AMPLITUDE EEG 18-27CPS (MICROVOLTS)
LEVEL OF INDUCED MUSCLE TENSION

FIG. 4. EEG measures as a function of induced muscle tension.

DISCUSSION

The physiological data were highly consistent in showing regular and continuous rise in level as a function of increments in tension induced in the right arm during tracking. Freeman's (1938) and Freeman and Simpson's (1938) results with palmar conductance were thus confirmed, but the additional data from muscle potentials, respiration, heart rate, and especially EEG considerably extend the possibilities of interpretation. The demonstration of such widespread changes, however, would appear to cast strong doubt on any rationale limited to the skeletal-motor (or somatic) sys-

tem. It appears that Meyer's (1953) interpretation suffers from this limitation. The results of this experiment definitely encourage the idea that the proprioceptive return from induced muscular tension produces widely generalized physiological effects. It therefore appears that Malmo (1959) was correct when he suggested that induced tension is one of the ways in which level of activation can be varied.

The EEG data require some special explanation. Generally, the terms "arousal" and "activation" have been used to refer to desynchronization in the EEG tracing in association with stimulation. This is what Lindsley (1951) refers to as "activation pattern" (p. 505). It is well known, however, that this reaction does not inevitably follow an alerting stimulus. The 8–12 cps component of the human EEG may, in fact, be augmented by an alerting stimulus if *S* is sufficiently drowsy. Stennett (1957) has systematically investigated the direction of the 8–12 cps change as a function of the *S*'s prior level of activation showing it to be an inverted U function.

In the present experiment, Stennett's (1957) practice of obtaining average voltage by means of band-pass filters was employed. This method permitted independent measurements of 8–12 cps EEG amplitude and 18–27 cps EEG amplitude. While the overall change in 8–12 cps EEG might have been expected to occur in the downward direction because of the generally activating effects of induced tension, finding a change in the upward direction was by no means unprecedented. In addition to Stennett's results there are recent (unpublished) findings of Malmo[4] showing a significant overall rise in 8–12 cps amplitude in going from a condition of lower, to one of higher, activation level. Malmo and Surwillo (1960) have also shown that under appropriate conditions an overall fall of 8–12 cps EEG can occur with a rise in activation level.

[4] R. B. Malmo, personal communication.

The evidence from Cond. EX (Table 1) quite clearly suggests an interaction between tension induction and tracking. That is, if induced tension were the sole factor, the points on the curves for Cond. EX should be more continuous with the points for Cond. VH than they in fact are. These results support the previous contention of Malmo (1959, p. 373) that activation level is invariably determined by an interaction between internal factors (mainly those affecting the "tonic baseline" activity of the ARAS) and external (cue) factors.

Performance per se in this experiment was also affected by induced muscle tension but not in the expected direction. If it is assumed that *S*s were relatively relaxed at the lower levels of muscle tension, then one would expect performance to be enhanced with induced muscle tension. On the contrary, however, there was a decrement in performance, and scores rose to prelearning levels. One possible explanation for this is that *S*s were overactivated at even the lowest tension level. However, the marked increase in physiological measures of activation with increased muscle tension above the Very Light condition would suggest that *S*s were quite relaxed at the lower levels. Another possible explanation is that the necessity of squeezing the dynamometer while tracking represented a sufficiently difficult task to require divided attention. For this explanation to be adequate, it must be supposed that the difficulty associated with maintaining tension increased with tension level. However, the strain gauge integrated output actually showed amount of tension to be constant over the length of a given trial, regardless of the amount of tension that was induced.

In seeking an explanation for the unfavorable effects of induced tension, the kind of tracking task selected for this experiment should be carefully considered. In tracking such a very slow moving target with its highly predictable pattern, movements are smooth and relaxed in contrast to the high tension movements required to do well in tracking a high frequency, irregular target

course under visual control. It seems quite possible that had the latter (high tension) kind of task been selected the results may have been different, with induced tension improving performance rather than degrading it.

Summary

Induced muscle tension has been demonstrated to have significant effects in relation to a wide range of behavioral phenomena. From the behavioral evidence, it appeared reasonable to consider that induced tension was only one of the many ways in which activation level can be varied. It followed that if induced tension were a reliable means of varying activation level, then in addition to the behavioral effects which have already been demonstrated, induced tension should also produce regular and consistent changes in the various physiological indicants of activation. The purpose of this experiment was to investigate such physiological changes.

Thirty-eight male college students were trained in an auditory tracking task. Physiological indicants of activation included heart rate, respiration rate, palmar conductance, frontal and occipital EEG, and EMGs from active and passive limbs. Following learning trials, *Ss* were required to track while squeezing a hand dynamometer in order to maintain one of five predetermined levels of muscle tension.

Results clearly showed close agreement between amount of tension induced and the level of activity in all physiological measures. These results were considered in support of a theory that the proprioceptive return from the induced muscle tension produces generalized behavioral and physiological effects indirectly by increasing activity in the reticular activating system.

References

Courts, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367.

Davis, J. F. Manual of surface electromyography. *USAF WADC tech. Rep.*, 1959, No. 59–184.

Davis, J. F., Stennett, R. G., & Quilter, R. E. An auditory tracking device designed for use in conjunction with continuous EEG recording. *Percept. mot. Skills*, 1957, **7**, 239–244.

Delafresnaye, J. F., Adrian, E. D., Bremer, F., & Jasper, H. H. (Eds.), *Brain mechanisms and consciousness.* Springfield, Ill.: Charles C Thomas, 1954.

Freeman, G. L. The optimal muscular tensions for various performances. *Amer. J. Psychol.*, 1938, **51**, 146–150.

Freeman, G. L., & Simpson, R. M. The effect of experimentally induced tension upon palmar skin resistance. *J. gen. Psychol.*, 1938, **18**, 319–326.

Jasper, H. H. Electroencephalography. In W. Penfield & T. C. Erickson (Eds.), *Epilepsy and cerebral localization.* Springfield, Ill.: Charles C Thomas, 1941. Pp. 380–454.

Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

Malmo, R. B. Activation: A neuropsychological dimension. *Psychol. Rev.*, 1959, **66**, 367–386.

Malmo, R. B., & Davis, J. F. A monopolar method for the measurement of palmar conductance. *Amer. J. Psychol.*, 1961, **74**, 106–113.

Malmo, R. B., & Surwillo, W. W. Sleep deprivation: Changes in performance and physiological indicants of activation. *Psychol. Monogr.*, 1960, **74**(15, Whole No. 502).

Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.

Ross, W. R., & Davis, J. F. Stable bandpass filters for electroencephalography. *IRE Canad. Conv. Rec.*, 1958, No. 860, 202–206.

Schnore, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *J. exp. Psychol.*, 1959, **58**, 117–128.

Stennett, R. G. The relationship of alpha amplitude to the level of palmar conductance. *EEG clin. Neurophysiol.*, 1957, **9**, 131–138.

(Received August 27, 1960)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 532–533

REPLICATION REPORT: TWO FAILURES TO REPRODUCE EFFECTS OF ANXIETY ON EYELID CONDITIONING

MARGARET S. KING, GREGORY A. KIMBLE
Duke University
JOHN GORMAN, AND RICHARD A. KING
University of North Carolina

Taylor (1951), Spence and Taylor (1951), Spence and Weyant (1960), and Baron and Connor (1960), among others, have shown that highly anxious *Ss* develop conditioned eyeblinks at a faster rate and to a higher asymptote than less anxious *Ss*. The two experiments reported here are practically identical in methodology with those mentioned above. They failed to show that highly anxious *Ss* condition better than less anxious *Ss*.

Experiment I[1]

Method.—This experiment investigated the relation between anxiety and the presence or absence of a ready signal in eyelid conditioning. The *Ss* were chosen from introductory psychology classes on the basis of extreme scores on the Manifest Anxiety (*MA*) scale. Those who obtained scores from 24–42 were designated as high-anxiety *Ss*; low-anxiety *Ss* were those who obtained scores from 1–7. One-half of the *Ss* in each of the anxiety groups received a verbal ready signal 2, 3, or 4 sec. prior to the CS light; the other *Ss* did not receive the ready signal. There were 16 *Ss*, 8 male and 8 female, in each of the four experimental groups. The apparatus, method of scoring CRs, and procedure were similar to that described by Dufort and Kimble (1958) and Spence and Taylor (1951).

Results.—The results are shown in Fig. 1. As can be seen, there is (*a*) no difference between the performance of the high-A and low-A groups which did not receive a ready signal; (*b*) a tendency for low-A *Ss* who received a ready signal to perform better than high-A *Ss* who received a ready signal (difference not significant on Trials 71–80); and (*c*) both high-A and low-A groups which received a ready signal performed at significantly lower levels than the no-ready groups ($F = 22.39$, $P < .001$, $df = 1/56$). This was determined by an analysis of variance in which ready signal, anxiety, and sex were the factors.

[1] This experiment was performed by Margaret A. King in partial fulfillment of the requirements for the MA degree at Duke University under the direction of G. A. Kimble.

The first finding, where there was no ready signal, is contrary to those of Spence and Weyant (1960) and Baron and Connor (1960). The second finding, where there was a ready signal, is contrary to those of Taylor (1951) and Spence and Taylor (1951). The third finding is in accord with one made by Dufort and Kimble (1958).

Experiment II[2]

Method.—This experiment investigated the relation between the *MA* scale measure of anxiety, the Welsh (1956) *A* measure of anxiety, and eyelid conditioning. The *Ss* were chosen from introductory psychology classes on the basis of high and low *MA* scale and Welsh *A* scores. On the *MA* scale, *Ss* having scores of 24–36 were high-anxiety *Ss*, while those having scores below 7 were low-anxiety *Ss*. On the Welsh *A* scale, *Ss* with scores above 22 were put in the high-anxiety group, while *Ss* with scores below 6 were put in the low-anxiety group. There were 15 male *Ss* in each of the *MA* scale groups (high *MA*

[2] This experiment was performed by John Gorman in partial fulfillment of the requirements for the MA degree at the University of North Carolina under the direction of R. A. King.

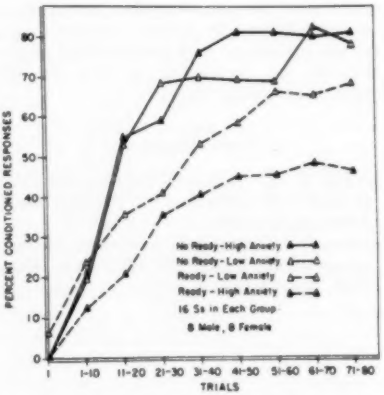

FIG. 1. Conditioning curves for the four groups of Exp. I.

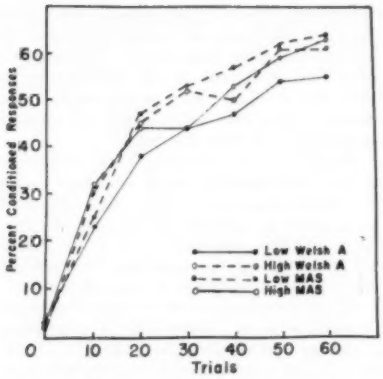

FIG. 2. Conditioning curves for the four groups of Exp. II.

scale, low *MA* scale) and 20 male *Ss* in each of the Welsh *A* groups (high Welsh *A*, low Welsh *A*). The apparatus, method of scoring CRs, and the procedure were similar to that described by Dufort and Kimble (1958) and Spence and Taylor (1951).

Results.—The results are shown in Fig. 2. As can be seen, there is (*a*) no significant difference between the high and low *MA* scale anxiety groups; and (*b*) no significant difference between the high and low Welsh *A* anxiety groups.

The contradictory findings presented here make it likely that anxiety as measured by the *MA* scale or the Welsh *A* scale is an interacting variable and has an effect upon conditioning only when it is combined with the proper values of other variables not explicitly manipulated by *E*.

REFERENCES

BARON, M. R., & CONNOR, J. P. Eyelid conditioned responses with various levels of anxiety. *J. exp. Psychol.*, 1960, **60**, 310–313.

DUFORT, R. H., & KIMBLE, G. A. Ready signals and the effect of interpolated UCS presentations in eyelid conditioning. *J. exp. Psychol.*, 1958, **56**, 1–7.

SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

SPENCE, K. W., & WEYANT, R. G. Conditioning performance of high and low anxious *Ss* in the absence of a warning signal. *J. exp. Psychol.*, 1960, **60**, 146–149.

TAYLOR, J. A. The relationship of anxiety to the conditioned response. *J. exp. Psychol.*, 1951, **42**, 81–92.

WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 264–281.

(Received February 25, 1961)

Journal of Experimental Psychology
1961, Vol. 62, No. 5, 533–534

SUPPLEMENTARY REPORT: EFFECTS OF INSTRUCTIONS ON THE PERCEPTION OF MULTIPLE TARGETS[1]

HARRY W. KARN AND LEE W. GREGG

Carnegie Institute of Technology

Karn and Gregg (1961) have previously reported a form of perceptual learning which required *S* to detect the presence or absence of a distinguishing binary cue, a dot, in each of three circles in a triangular array. Two of the three circles provided unchanging cue values in that the dot was always present in one location and never present in another. For the third circle, presence or absence of the dot was random. As a function of amount of experience with the complex at exposures of 1 sec.—ample time for *S* to observe without error the state of the three circles—significant error reduction occurred during test series when the exposure time of .1 sec. was too short to permit scanning of the entire stimulus field.

It was anticipated that *Ss* would rationally attack the perceptual task by learning the nature of the unchanging elements of the complex and then direct attention toward the changing element. Nearly perfect performance should ensue. Instead, the learning which occurred was slow and incremental and appeared to be the result of some kind of set. If cognitive sets play a critical role, performance should be dependent upon instructions given *S* concerning the task requirements. A simple suggestion or hint can result in insightful solutions in certain instances (Luchins, 1942; Maier, 1933).

As an initial attempt to provide a better understanding of the behavior generated by

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-770.

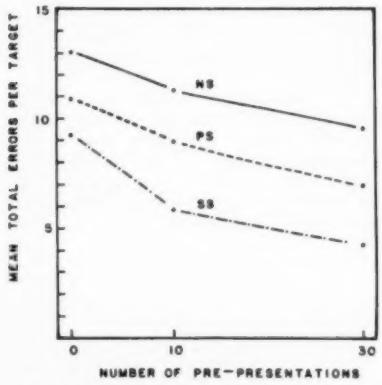

FIG. 1. Mean total errors per target as a function of number of prepresentations at 1 sec. for nonspecific (NS), problem solving set (PS), and specific solution groups (SS).

the multiple stimulus task the instructions were modified to induce a problem solving set at two levels of explicitness.

Method.—The apparatus, general procedure, and group allocation of *Ss* were the same as those of the previous study (Karn & Gregg, 1961). Three configurations (Cs) defining the specific combinations of loading (presence, absence, or random occurrence of a dot in a circle) and location in the triangular array were used. These Cs form a latin square so that each loading occurs once and only once at each location. Amount of prior experience was varied by presenting 30, 10, or 0 prepresentations of the particular C at 1 sec. exposure. There followed 60 presentations of the same C at .1 sec.

The first set of instructions were nonspecific (NS) in that *S* was informed that sometimes one or more of the circles would contain a dot. A report on all circles was required.

To induce a problem solving set (PS) the following was added to the basic instructions. "Even though you cannot see all of the circles at a single presentation, it is possible to get all of them correct. Try to discover how this can be done."

The third set of instructions were worded in such a way that a specific solution could be inferred (SS): "Even though you cannot see all of the circles at a single presentation, you can get all of them correct by keeping in mind that only one of the circles changes from trial to trial. Two of the circles never change. Use this information to improve the accuracy of your reporting."

Results and discussion.—Total errors for each of the separate target locations during 60 presentations at .1 sec. were obtained. An analysis of variance performed for these data yielded significant ($P < .01$) main effects for Location, Loading, Instructions, and Number of prepresentations at 1 sec.

Figure 1 presents the error reduction curves for each level of instruction as a function of the number of prepresentations at 1 sec. The effect of instructions is to lower the overall magnitude of errors. Differences in the slopes of these curves are not significant; rate of error reduction remains essentially constant.

The failure to obtain insightful shifts in performance suggests that the perceptual processing which occurs remains at the level of visual discrimination based on criteria as yet unspecified. The lack of evidence for a rational processing of the information inherent in the stimulus complex as a function of problem solving sets, does not mean that learning sets should be discounted or minimized in explaining the behavior. A variety of unique learning sets have been suggested or implied as playing critical roles in perceptual learning situations. Nonverbalized perceptual and motor sets are possibilities as are related eye movement biases in search tasks. The precise ways in which these factors govern the behavior under investigation remains to be determined. Whatever their roles, they appear to have a hampering influence on the problem solving process envisaged in terms of a selective utilization of information provided by the stimulus complex.

REFERENCES

KARN, H. W., & GREGG, L. W. Acquisition of perceptual responses as a function of loading, location, and repetition. *J. exp. Psychol.*, 1961, **62**, 62–69.

LUCHINS, A. S. Mechanization in problem solving. *Psychol. Monogr.*, 1942, **54**(6, Whole no. 248).

MAIER, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, **24**, 144–155.

(Received March 13, 1961)

Effective *January 1, 1962,* all **regular** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant; Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. **Early publication** articles should be sent to the Editor: Arthur W. Melton; Department of Psychology, University of Michigan; Ann Arbor, Michigan until *July 1, 1962.* Thereafter they should be sent to the Editor-elect.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.